Synthese (2021) 198:6889–6916
https://doi.org/10.1007/s11229-019-02496-z

NORMS FOR RISK

Deontological decision theory and lesser-evil options

Seth Lazar[1] · Peter A. Graham[2]

Received: 14 April 2019 / Accepted: 26 November 2019 / Published online: 5 December 2019

Abstract
Normative ethical theories owe us an account of how to evaluate decisions under risk and uncertainty. Deontologists seem at a disadvantage here: our best decision theories seem tailor-made for consequentialism. For example, decision theory enjoins us to always perform our best option; deontology is more permissive. In this paper, we discuss and defend the idea that, when some pro-tanto wrongful act is all-things considered permissible, because it is a 'lesser evil', it is often merely permissible, by the lights of deontology. We show that this raises new problems for deontological decision theory, and we show that to resolve them, we need to take a more innovative approach to morally evaluating decision-making under risk and uncertainty.

Keywords Risk · Decision theory · Lesser-evil options · Parity · Harm

1 Introduction

The moral justification for every decision we make depends on non-moral facts that are almost always in doubt. If a moral theory does not apply to risky choices, then it is crucially incomplete. Deontologists have only recently begun to account for risk and uncertainty.[1] One promising approach uses existing resources in decision theory

[1] We follow standard decision-theoretic practice by using 'decision-making under risk' to refer to choices with sharp probabilities, and 'decision-making under uncertainty' to refer to choices with-

Seth Lazar presented this paper at the 'Ethics and Risk' Workshop at the ANU, and at a workshop on Ethics and Uncertainty at Hebrew University of Jerusalem. Thanks to Renée Bolinger and Ittay Nissan-Rozen for organising those incredibly productive meetings. For written comments, special thanks to Adam Bales and Alan Hájek, who each commented on two different drafts. Thanks also to Claire Benn, Julia Staffel, and Orri Stefánsson for further helpful comments. Research for this paper was supported by ARC Grant DP170101394. Lazar and Graham are equal co-authors of this paper.

✉ Seth Lazar
seth.lazar@anu.edu.au

[1] School of Philosophy, Australian National University, Office 3227 Coombs Building 9, Acton, ACT 2601, Australia

[2] University of Massachusetts, Amherst, Amherst, USA

to offer deontologists a 'criterion of subjective permissibility'—a set of necessary and sufficient conditions for an act's being morally permissible in light of one's epistemic limitations.[2]

Some, of course, scoff at this combination.[3] Decision theory and *consequentialism* look like natural partners.[4] Why would anyone try to develop a deontological decision theory? And yet, deontologists have focused almost exclusively on developing objective moral theories, while decision theorists are experts in conceptualising decision-making under risk and uncertainty. When seeking to extend our deontological moral theories to this uncharted domain, it would be hubris to simply ignore the experts. Of course, decision theory is no less disputatious a field than moral philosophy. There is, though, a simple, classical form of decision theory, of the kind that Bernoulli or Pascal had in mind. And that simple decision theory's commitments are actually rather modest—and rather compelling. Deontologists' knee-jerk scepticism about a partnership with this kind of view may perhaps be overcome.[5]

The big question, then, is whether there are any *genuinely* insurmountable obstacles to a successful partnership between deontology and decision theory. If not, so much the better for deontologists. But if there are, then we must either push the boundaries of decision theory itself or choose a different path altogether.

One central commitment of decision theory, construed this way, is anathema to deontology: the assumption that we can adequately represent any rational agent's choices with a single utility function, ranking outcomes along one dimension.[6] Because of this commitment, a simple version of deontological decision theory

Footnote 1 (continued)

out sharp probabilities. For the recent deontological literature, see e.g.: Seth Lazar, 'In Dubious Battle: Uncertainty and the Ethics of Killing', *Philosophical Studies* 175/4 (2018), 859–883; Seth Lazar, 'Deontological Decision Theory and Agent-Centered Options', *Ethics* 127/3 (2017), 579–609; Seth Lazar, 'Anton's Game: Deontological Decision Theory for an Iterated Decision Problem', *Utilitas* 29/1 (2017), 88–109; Horacio Spector, 'Decisional Nonconsequentialism and the Risk Sensitivity of Obligation', *Social Philosophy and Policy* 32/2 (2016), 91–128; Sergio Tenenbaum, 'Action, Deontology, and Risk: Against the Multiplicative Model', *Ethics* 127/3 (2017), 674–707. Some consequentialists have offered different approaches, e.g. Mark Colyvan, Damian Cox, and Katie Steele, 'Modelling the Moral Dimension of Decisions', *Noûs* 44/3 (2010), 503–529; Yoaav Isaacs, 'Duty and Knowledge', *Philosophical Perspectives* 28/1 (2014), 95–110.

[2] One of the authors is sceptical about the concept of subjective permissibility, so construes our enterprise as formulating a theory of how a morally conscientious person would act under uncertainty. Note that we use 'subjective permissibility' as an umbrella term to encompass any account of permissibility under limited information. Everything we say could apply equally well to what Parfit called 'belief-relative' and 'evidence-relative' permissibility. We could also interpolate more exotic interpretations of probability.

[3] Isaacs, 'Duty'. Sergio Tenenbaum doesn't scoff, but he is sceptical: Tenenbaum, 'Action, Deontology, and Risk'.

[4] Indeed, some have argued that the necessity of developing a deontological decision theory suggests that deontology itself is, in fact, just a subspecies of consequentialism. Graham Oddie and Peter Milne, 'Act and Value: Expectation and the Representability of Moral Theories', *Theoria* 57/1–2 (1991), 42–76; Rachael Briggs, 'The Anatomy of the Big Bad Bug', *Noûs* 43/3 (2009), 428–449; Douglas W. Portmore, 'Uncertainty, Indeterminacy, and Agent-Centred Constraints', *Australasian Journal of Philosophy* 95/2 (2017), 284–298. We do not agree with this view.

[5] Colyvan et al., 'Modelling'; Lazar, 'Agent-Centred Options'; Oddie and Milne, 'Act and Value'.

[6] Lazar, 'Agent-Centered Options'.

cannot adequately accommodate *agent-centred options to act suboptimally*—that is, the moral licence we have to either prefer or sacrifice our own interests, even when doing so is not overall morally best. However, we can resolve this problem without too seriously departing from simple decision theory, by adapting our decision rule. Instead of saying that rational agents must *maximise* expected moral utility, we can build in a licence to forbear from doing so, either when the personal costs are excessive or when the best act is best only in virtue of personal benefits.[7] In effect, we solve the problem by shifting from comparing acts along one dimension (roughly, moral weight) to comparing them along two (moral weight and personal good).

In this paper, we introduce a further deontological commitment, which is, in our view, likely to be endorsed by many deontological moral philosophers when they reflect on it.[8] This is the view that, in cases where one is permitted to harm some for the sake of others, because doing so is the lesser evil, one is very often merely permitted to do so, rather than required. We will call these ‘lesser-evil options’.

Consider the most famous case in deontological ethics: a trolley will kill five unless diverted onto a side-track, where it will kill one.[9] You are at the lever, and can decide who lives or dies. Many deontologists think that pulling the lever is permissible. Most of those, we conjecture, will deny that doing so is required. If killing the one is permissible, then this (as many deontologists would have it) is because it is the lesser of two evils—killing one is bad, but letting five die is worse. So if you are merely permitted, not required, to turn the trolley, then you have a ‘lesser-evil option’.

Since lesser-evil options are more often presupposed than discussed—indeed, we are aware of only one, moderately sceptical, paper on the topic—part of our task in this paper will be to introduce them, and defend their prima facie plausibility.[10] But our central goal is to discover whether deontological decision theory can adequately accommodate lesser-evil options. We have already introduced an additional dimension of normative strength in order to accommodate agent-centred options. Must we introduce yet another to cater for lesser-evil options? Or can one kind of option be reduced to the other—or else both to some further, deeper thing? Or does the existence of lesser-evil options pose an insurmountable obstacle to developing a deontological decision theory? Those are our questions in this paper.

[7] Dual-ranking versions of act consequentialism have been developed by Theodore Sider, ‘Asymmetry and Self-Sacrifice’, *Philosophical Studies* 70/2 (1993), 117–132; Jean-Paul Vessel, ‘Supererogation for Utilitarianism’, *American Philosophical Quarterly* 47/4 (2010), 299–319; Douglas W. Portmore, *Commonsense Consequentialism: Wherein Morality Meets Rationality* (Oxford: Oxford University Press, 2011).

[8] Helen Frowe, ‘Lesser-Evil Justifications for Harming: Why We’re Required to Turn the Trolley’, *The Philosophical Quarterly* 68/272 (2018), 460–480.

[9] Philippa Foot, *Virtues and Vices, and Other Essays in Moral Philosophy* (New York: Oxford University Press, 2002); Judith Jarvis Thomson, *Rights, Restitution, and Risk: Essays in Moral Theory* (Cambridge, Mass.: Harvard University Press, 1986).

[10] We discuss Frowe, ‘Lesser-Evil Justifications for Harming: Why We’re Required to Turn the Trolley’ in a footnote to Sect. 4 below. Note that Frowe’s view vindicates the existence of lesser-evil options in general, though she denies that they apply to standard trolley cases.

In the next section, we'll give more background, explain what we mean by deontology, why deontology needs a decision theory, and discuss what deontological decision theory might look like. In Sect. 3, we'll introduce lesser-evil options, and show that *if* we have lesser-evil options, then existing deontological decision theories are inadequate. Before trying to develop an alternative, though, we ask whether the antecedent of that conditional is satisfied. In Sect. 4 we tentatively argue that it is. Lesser-evil options are well motivated, and naturally cohere with some central tenets of deontological ethics. In Sect. 5 we'll develop a new criterion of subjective permissibility which can accommodate lesser-evil options under risk, while departing from classical decision theory, and we'll point the way to future research.

2 Introducing deontological decision theory

If we're going to talk about the union of deontology and decision theory, we need to say a little about how we understand each term. This is especially important since neither decision theorists nor deontologists have an established record of engaging with one another's work, and given how internally diverse both camps are. Setting some parameters at the outset will help pinpoint, for each side, which version of the other side is on the table. First, deontology.

2.1 What we mean by deontology

We care about deontology's *deontic verdicts*—that is, its judgements of requirement, permissibility and impermissibility—and the reasons it gives for reaching them. We are not interested, here at least, in deontological metaethics. Our kind of deontologists think that humans (and perhaps some other animals) have moral status, probably grounded in our capacities for rationality and moral reasoning. Because of that status, we are owed—and owe one another—respect. In particular, beings with moral status have duties to one another, grounded in that shared moral status. These duties are not all equally important or demanding. Deontologists typically think that duties not to harm (negative duties) outweigh duties to benefit (positive duties). And, other things equal, one can be required to bear more cost to abide by a negative duty than to perform a positive duty. Some duties are 'directed', insofar as they are owed to someone in particular, some 'undirected', insofar as you have a duty to ϕ without owing it to anyone in particular.

There are hierarchies of weight and stringency within classes of duty. These hierarchies are not simply determined by the consequences of non-compliance. For example, it is harder to justify intentionally, knowingly, maliciously breaching one's duty not to harm another, as a means to realise some personal benefit, than it is to justify breaching that duty unintentionally, unwittingly, without malice, and not as a means.

Breaching a directed duty involves incurring a kind of moral debt, which can be repaid only by apology, compensation or some form of symbolic reparation. This is true even when breaching that duty is the only way to avoid an even worse outcome. Even in such cases of *lesser-evil justification*, there is always a moral remainder.

Sometimes, however, an act that would otherwise be a breach of duty is not, because something the target has done has made her liable to be harmed in that way.[11] For example, if one person culpably attacks another, knowingly risking her life, then the attacker can be liable to be killed by her would-be victim, in self-defence.

Deontologists typically believe that you should prioritise your own duty-fulfilment above seeing to it that others fulfil their duties. Many explain this by arguing that our duties give us *agent-relative reasons*—reasons that apply only or with special force to the agent[12]; others use different explanatory concepts.[13] All deontologists would agree that you ought not to breach your duty X, even if doing so would prevent otherwise identical breaches of X by two other people. Duties may also give *agent-neutral* reasons, so we may have some reason to see to it that others abide by their duties. But this is weaker than our grounds for fulfilling our own.

Thus far we have focused on the special constraints that bind deontologists—the many considerations that can make an act impermissible, and determine how seriously wrong an impermissible act is. But deontology is not only about constraint. Positive permissions are equally central. Deontologists typically think that we are sometimes permitted not to choose our morally best alternatives, on any sensible understanding of what 'best' means. You are not required to choose an option that is unambiguously overall morally better than all the other alternatives, if it involves unreasonable personal costs. And you might be permitted to forgo a morally better alternative, when it is better only in virtue of the fact that it benefits you.[14]

While deontological ethics has made huge progress over the last four decades, it still has one glaring blind spot.[15] Deontologists have failed either to systematically explain how to apply their moral theories to situations with imperfect information or else to provide an argument to justify this omission. Our goal is to explore whether decision theory can help us extend a deontological view like this to the context of risk and uncertainty.

2.2 Deontologists need a criterion of subjective permissibility

Most work in deontology focuses exclusively on objective permissibility. Those who have addressed risk and uncertainty have done so piecemeal, offering principles

[11] See e.g. Jeff McMahan, 'The Basis of Moral Liability to Defensive Killing', *Philosophical Issues* 15/1 (2005), 386–405.

[12] Philip Pettit and Robert Goodin, 'The Possibility of Special Duties', *Canadian Journal of Philosophy* 16/4 (1986), 651–676; David McNaughton and Piers Rawling, 'Agent-Relativity and the Doing-Happening Distinction', *Philosophical Studies* 63/2 (1991), 167–185.

[13] F. M. Kamm, 'Review: Non-Consequentialism, the Person as an End-in-Itself, and the Significance of Status', *Philosophy and Public Affairs* 21/4 (1992), 354–389; Victor Tadros, *The Ends of Harm: The Moral Foundations of Criminal Law* (Oxford: Oxford University Press, 2011); Tom Dougherty, 'Agent-Neutral Deontology', *Philosophical Studies* 163/2 (2013), 527–537. See Matthew Hammerton, 'Is Agent-Neutral Deontology Possible?', *Journal of Ethics and Social Philosophy* 12/3 (2017), 319–324 for whether deontology is necessarily agent-relativist.

[14] There is a huge literature on agent-centred options. For an overview and bibliography, see Seth Lazar, 'Moral Status and Agent-Centred Options', *Utilitas* 31/1 (2019), 83–105.

[15] Many have made this observation: see e.g. Barbara H. Fried, 'What Does Matter? The Case for Killing the Trolley Problem (or Letting It Die)', *Philosophical Quarterly* 62/248 (2012), 505–529.

suited only to narrow domains, such as beneficence, self-defence or war. But any domain-specific approach to decision-making with imperfect information faces the obvious question: what should we do when we don't know which domain is salient? Given the infinite variety of possible 'domain-mixing' scenarios, we clearly need some systematic principle that considers all possibilities.[16] We need a criterion of subjective permissibility—an account that picks out which acts are required, permissible and wrong, given the agent's uncertainty.[17]

Of course, this is not all we need. We also need an account of how imperfect information can directly affect our *objective* moral reasons.[18] And clearly we should aim to develop heuristics and decision procedures that people can actually deploy, in order to make good choices under risk and uncertainty. This is not a task for philosophers alone. Developing actual deliberation procedures to be applied in practice requires insights from many other academic fields, as well as the practical wisdom borne of experience. For example, nobody should let a moral philosopher write, on her own, rules of engagement that fit on a credit card, for soldiers to take into battle.

Since all decision-making occurs without access to the relevant non-moral facts, we think that any successful moral theory should tell us how to rank acts, and which are permissible, in light of that limited epistemic position. Even though such a theory would not be easy to apply in many actual cases, it would be action-guiding. It can consider the agent's feasible set of actions under the description by which they appear to her, rank them, and say which are permissible. A criterion of objective permissibility fails to be action-guiding even in this sense. Moreover, if there are cases in which the permissibility of one's action depends on what would have happened if one had done otherwise, and there is no fact of the matter about what that would have been, then a criterion of subjective permissibility may be the *only* way to determinately morally assess one's action.[19] A criterion of subjective permissibility can also offer guidance on which acts count as *robustly* morally permissible, without being dependent on contingent facts, beyond a person's control, about how the world actually turns out.

So, while deontologists awakening to the reality of risk and uncertainty have much work to do, developing a criterion of subjective permissibility is an important

[16] As Alan Hájek has pointed out to us in correspondence, the domain-based approach may face a version of the 'reference-class problem', when a given choice is in multiple domains. There may also be problems evaluating sequences of choices, when the relevant domain shifts over the course of the sequence.

[17] We are at near-opposite ends of the debate between *objectivists* and *subjectivists* about morality: one of us is an all-out objectivist, the other wavers between the 'sense-splitting view' and all-out subjectivism. While we agree on the importance of developing something that one of us describes as a criterion of subjective permissibility, the other would describe it as an account of a distinct, non-moral ought. See Peter A. Graham, 'In Defense of Objectivism About Moral Obligation', *Ethics* 121/1 (2010), 88–115.

[18] E.g. Seth Lazar, 'Risky Killing and the Ethics of War', *Ethics* 126/1 (2015), 91–117; Patrick Tomlin, 'Subjective Proportionality', *Ethics* 129/2 (2019), 254–283.

[19] Oddie and Milne, 'Act and Value'; Portmore, 'Uncertainty, Indeterminacy'.

first step. And the natural way to begin that project is by using the existing tools of decision theory.[20]

2.3 Introducing 'classical' decision theory

There is a simple, classical version of decision theory which offers an obvious point of departure for deontological ethics. When choosing under imperfect information, identify your available acts, and the different ways the world might be (states). Then, for each act-state pair, use numbers to represent how well supported by reasons that act would be, if that state were actual, and how likely that state is to be actual, if you choose that act.[21] Multiply those numbers together, sum them for each act across all possible states, and you have a measure of their support by your probability-weighted reasons. Perform the act best supported by your probability-weighted reasons.[22]

We call this 'classical' decision theory since—if we change the terminology a little—it is the understanding of decision theory as it was first developed by, for example, Pascal, Bernoulli and Keynes, and to some extent Ramsey.[23] However, after von Neumann and Morgenstern showed (building on Ramsey's work) that any agent whose preferences over gambles obeyed some seemingly innocuous axioms could be *represented* as maximising expected utility, much subsequent work in philosophy and economics focused less on applying classical decision theory to actual choices, and more on the mathematics that underpins those representation theorems.[24] Some now adopt a 'constructivist' approach to decision theory, according to which probability and utility functions are no more than constructs generated by one's preferences over gambles—the only real constraints on rational choice are given by the

[20] Michael J. Zimmerman, *Living with Uncertainty: The Moral Significance of Ignorance* (Cambridge: Cambridge University Press, 2008); Colyvan et al., 'Modelling'; Oddie and Milne, 'Act and Value'; John Broome, *Weighing Goods: Equality, Uncertainty and Time* (Oxford: Blackwell, 1991); Spector, 'Decisional Nonconsequentialism and the Risk Sensitivity of Obligation'.

[21] The numbers representing likelihoods should be probabilities, in the interval [0, 1], such that the sum of the probabilities of every state equals 1.

[22] Note that the terms we use here are chosen for their palatability to deontologists; classical decision theory expressed the same ideas in different terms. Thanks, throughout this section, to Alan Hájek, for extensive comments and discussion.

[23] Alan Hájek, 'Interpretations of Probability', (Winter 201 edn.); Antoine Arnauld, *Logic, or, the Art of Thinking ("the Port-Royal Logic")* (Indianapolis: Bobbs-Merrill); Daniel Bernoulli, '"Specimen Theoriae Novae De Mensura Sortis", Commentarii Academiae Scientiarum Imperialis Petropolitanae, 5: 175–192. English Translation, 1954, "Exposition of a New Theory on the Measurement of Risk"', *Econometrica* 22/1 (1954), 23–36; John Maynard Keynes, *A Treatise on Probability* (London: Macmillan, 1921). Note that Keynes termed this taking the 'mathematical expectation' of an act; he was not wholly on board with it as an approach to decision-making (wanting also to emphasise other properties of evidence, such as its 'weight'), but he clearly understood this to be the default approach to decision-making under risk. Ramsey sits at the cusp between classical and constructivist decision theory; both sides can claim him as their own with some credibility.

[24] John von Neumann and Oskar Morgenstern, *Theory of Games and Economic Behavior* (Princeton: Princeton University Press, 1944); Leonard J. Savage, *The Foundations of Statistics* (New York: Wiley, 1954).

formal rationality axioms.[25] In recent years, some decision theorists have pushed back against this approach, questioning the fruitfulness of focusing on representation theorems.[26]

It seems to us that the purpose of representation theorems is to *justify* a certain approach to decision-making under risk, by showing that it is implied by some plausible constraints on rational choice. If we then apply that approach to actual choices, and find that on any plausible probability or utility function it implies intolerable results, then that surely gives us reason to question those putative constraints.[27] The decision rule implied by even von Neumann and Morgenstern's decision theory would seem to be fair game.

Irrespective of how that debate proceeds, constructivist decision theory will be of little help to moral philosophers, since it takes preferences over gambles as given, and our task is precisely to work out what our preferences over (moral) gambles ought to be. Obedience to the rationality axioms guarantees only a certain kind of internal coherence or consistency—and it is clearly quite possible to be a coherent and consistent moral monster. What's more, the basic insights of classical decision theory have proven their mettle in many different practical applications, from risk-management in general, to insurance and other branches of actuarial science, high-frequency algorithmic trading, robotics and AI. These many practical successes suggest that classical decision theory, understood in the very minimal and non-committal way described above, should be the default starting point for moral philosophers approaching decision-making under risk.[28]

And yet many deontologists will baulk at this. We've used different terms, but a rose by any other name would smell as consequentialist … Classical decision theory tells us to 'maximise expected utility'. Isn't that just consequentialism?

Well, not really. We chose our terms carefully. Decision theory *does* look a lot like consequentialism, if we talk about ranking outcomes with respect to their utility. But

[25] For two good overviews, see Lara Buchak, 'Decision Theory', in Alan Hájek and Christopher Hitchcock (eds.), *Oxford Handbook of the Philosophy of Probability* (Oxford: Oxford University Press, 2016); Rachael Briggs, 'Normative Theories of Rational Choice: Expected Utility', in Edward Zalta (ed.), *Stanford Encyclopaedia of Philosophy*.

[26] E.g. Christopher J. G. Meacham and Jonathan Weisberg, 'Representation Theorems and the Foundations of Decision Theory', *Australasian Journal of Philosophy* 89/4 (2011), 641–663; Kenny Easwaran, 'Decision Theory without Representation Theorems', *Philosophers' Imprint* 14/27 (2014), 1–30; Martin Peterson, 'An Argument for the Principle of Maximizing Expected Utility', *Theoria* 68/2 (2002), 112–128. As Al Hájek points out in his MS 'Risky Business', these philosophers, and Hájek himself, are picking up on themes in Richard Jeffrey's influential work. Jeffrey argued that a number of these supposed axioms of rationality were primarily shaped by mathematical convenience: the impartiality axiom 'is not the sort of assumption that is particularly plausible simply because we are taking prospects to be propositions. The axiom is there because we need it, and it is justified by our antecedent belief in the plausibility of the result we mean to deduce from it', Richard C. Jeffrey, *The Logic of Decision* (Chicago and London: The University of Chicago Press, 1983): 147. For other criticisms of treating expected utilities as more fundamental than probabilities and utilities themselves (as the representation theorem approach does), see Lina Eriksson and Alan Hájek, 'What Are Degrees of Belief?', *Studia Logica* 86/2 (2007), 183–213; Alan Hájek, 'Arguments for—or against—Probabilism?', *The British Journal for the Philosophy of Science* 59/4 (2008), 793–819.

[27] See, for example, Lara Buchak, *Risk and Rationality* (Oxford: Oxford University Press, 2013).

[28] Thanks to a reviewer for pressing us on this.

an outcome is really just an act–state pair. And ‘utility’, understood most charitably, is just a measure for degree of ‘rational support’.[29] And deontologists can surely discuss the level of rational support there is for a particular act, given that a particular state is the case. This idea of ‘rational support’ can include appeal to both agent-relative and agent-neutral reasons, and indeed all the other distinctions that deontological ethics recognises. We find it hard to imagine what the task of moral philosophy would be, if not delineating our moral reasons for action, given that the world is thus and so.

In this simple, classical form, decision theory involves very few actual commitments. These three are, we think, fundamental:

(1) Reasons have weight that can be quantitatively represented;
(2) When choosing under risk, we should discount the net reasons for an act, given a state, in linear proportion to the probability of that state being actual, given that act;
(3) We should perform the act that is best supported by our probability-weighted reasons.[30]

These may not be innocent commitments. But they are not inherently opposed to deontology.

Numbers might make for an artificial representational device, but deontologists can certainly represent the weights of reasons in this way (and commonly do). The practice of representing reasons with numbers does not imply, for example, that they are infinitely precise, or ‘mathematically well-behaved’.[31]

Perhaps one might reject (2), on the grounds that one favours a non-neutral attitude to risk.[32] But while morality may well mandate *some* attitude to risk, there doesn’t seem to be any reason to think that *deontologists* in particular should favour one risk attitude or another. Risk-neutrality is a reasonable starting point; we can of course argue for a departure from it, but one’s views on the morally appropriate attitude to risk are not obviously predetermined by one’s commitment to deontology.

Point (3) will prove to be a real problem. However sophisticated your interpretation of the deontological moral utility function, any decision rule with the structure

[29] Obviously, decision theory did not arrive at this broad, encompassing notion of utility directly. Pascal was aiming to maximise financially; Bernoulli introduced the idea that we should care about utility rather than money per se. Arguably, the understanding of utility as such a capacious formal construct came about only in the mid-twentieth century. Thanks again to Alan Hájek here.

[30] Notice that even constructivist decision theory is committed to a version of these three views, which says that rational agents can be represented in a way that satisfies 1–3.

[31] Note that numbers are also an artificial way of representing degrees of belief. Thanks again to Alan Hájek.

[32] Buchak, *Risk and Rationality*.

'maximise expected moral utility' will ultimately fail.[33] There is a genuine fault-line between deontology and classical decision theory.[34]

Fortunately, however, there are at least two ways to retain the core insights of classical decision theory while accommodating agent-centred options. One could either reject maximising in favour of satisficing, or one could adopt additional constraints within one's decision rule. Others have discussed the problems with satisficing[35]; perhaps most important here is that, however plausible it is as an interpretation of how people actually make decisions, it is a technical fix for representing agent-centred options, lacking any real motivation from within deontological ethics. A better approach is to figure out what grounds our options to act suboptimally, and build that into the decision rule. Here is one such rule:

COST: An act is subjectively permissible for an agent if and only if:

(a) there is no all-things-considered expectedly better act or
(b) every all-things-considered expectedly better act either

    (i) involves unreasonable marginal expected costs to the agent or
    (ii) is better only in virtue of expected benefits to the agent.[36]

One act is expectedly better than another just in case it is better supported by the agent's probability-weighted reasons for action, as determined in the way just outlined.[37] 'Marginal' in clause (i) is used in the economists' sense: consider a choice between ϕ and ψ, where ϕ is morally expectedly better than ψ. (i) is satisfied when ϕ is not expectedly better than ψ by enough to make the difference in expected personal cost reasonable. Note that 'marginal' is unnecessary in clause (ii) because it is already explicit about comparing the difference between options.[38] Some who are otherwise attracted to COST might deny that marginal costs are what matters in clause b(i)—that's fine, for their purposes assume that in the rest of the paper we are discussing COST', from which that word is removed.

[33] Lazar, 'Agent-Centered Options'.

[34] This fault-line is not recognised by the few philosophers who have sought to develop deontological decision theories modelled on classical decision theory, in particular Oddie and Milne, 'Act and Value'; Colyvan et al., 'Modelling'; Spector, 'Decisional Nonconsequentialism and the Risk Sensitivity of Obligation'; Kristian Olsen, 'Subjective Rightness and Minimizing Expected Objective Wrongness', *Pacific Philosophical Quarterly* 99/3 (2018), 417–441. Although their proposed principles differ in various respects, they all involve ranking acts along a single dimension, so all fail to adequately accommodate agent-centred options.

[35] For an overview and some novel objections, see Lazar, 'Agent-Centered Options'.

[36] Ibid.

[37] COST can be adapted to accommodate any interpretation of probability, though moral philosophers are generally most interested in something like evidential probabilities, or the subjective probabilities one would have, if one were a reasonable person, who did the morally appropriate research. As a referee has pointed out to us, there may be further reasons for constraining which interpretation of probability to deploy, grounded in decision theory rather than deontology.

[38] Thanks to Alan Hájek for pressing us on this.

The other details of COST need not detain us here. The crucial point is that, to accommodate agent-centred options, we need an additional way of comparing acts against one another. As well as comparing acts with respect to their level of support by probability-weighted moral reasons, we must also compare their costs to the agent, to determine whether those probability-weighted moral reasons are such as to make those marginal costs reasonable ones for the agent to bear. What makes a cost reasonable is as fundamental a question, in moral philosophy, as what makes one act morally better than another. For deontologists, we have to ask both questions—we can't read off an answer to one from our answer to the other. An act's being expectedly best is sufficient, but not necessary, for it to be permissible. It might also be permissible if the expectedly better alternatives either involve unreasonable marginal expected costs to the agent or are better only in virtue of expected benefits to the agent.

COST is not the only plausible adaptation of deontological decision theory. But any alternatives would have to share this crucial feature with COST: as well as ranking acts for their overall degree of moral support, we must also take costs to the agent into account. Some forms of 'dual-ranking act consequentialism' (DRAC), such as those defended by Douglas Portmore, could in principle be extended to decision-making under risk in a similar way.[39]

The move from classical decision theory to COST is not trivial. One of the primary justifications for maximising expected utility is that we can prove that any agent whose preferences over gambles obey some seemingly innocuous axioms can be represented as an expected utility maximiser, with a unique probability function, and a utility function that is unique up to positive affine transformation.[40] That utility function is one-dimensional. You can't adequately represent choices grounded in a two-dimensional utility function with a one-dimensional one. For example, deontologists who apply COST might be represented, if their 'moral preferences' over gambles were taken as inputs, as violating transitivity.[41]

We do not think this is a problem. We think that if morality *required* agents to act in ways that violate axioms of rationality, that would be unsettling (though perhaps not irremediably so). But COST merely *permits* preferences over gambles that do not abide by all the axioms of rationality. It does not require them. One *may*, according to COST, conform to the axioms of rationality in every choice. But one is not morally required to do so. We see no problem with this result. Indeed, COST explicitly states that one does no wrong by failing to advance one's own good (in clause b(ii)). So it already contends that some acts that are rational in this other sense—of advancing one's own good—are not morally required.

However, perhaps COST does not go far enough. As we will now argue, it seemingly fails to accommodate some other plausible options to act suboptimally.

[39] At least, DRAC can adequately emulate clauses (a) and (b)(i) of COST. It cannot adequately accommodate agent-sacrificing options. See Lazar, 'Accommodating Options'.

[40] Von Neumann and Morgenstern, *Theory of Games and Economic Behavior*.

[41] See F. M. Kamm, 'Supererogation and Obligation', *Journal of Philosophy* 82/3 (1985), 118–138. Others have noted this implication of adopting a two-dimensional approach, see e.g. Sider, 'Asymmetry'; Portmore, *Commonsense*.

3 Lesser-evil options, and error theories for them

Lesser-evil options arise when it is morally permissible, but not required, to perform the lesser of two evils. For example, under ordinary circumstances, you are permitted, but not required, to turn the trolley towards one, when there is no other way to save five. In this section, we will introduce lesser-evil options, show how one *could* try to explain them while still endorsing a principle like COST, and then argue that this attempt fails. We are forced to choose between denying that lesser-evil options exist and adapting our deontological decision theory further to accommodate them. In Sect. 4 we give reasons against the first approach; in Sect. 5 we attempt the second.

We'll characterise paradigmatic instances of lesser-evil options as follows:

> X is a lesser-evil option for agent A if and only if X involves the merely permissible (i.e. not required) imposition of harm H by A on some person or persons B_{1-n}, who is/are not liable to suffer H, because every alternative is all-things-considered morally worse.

Suppose that we grant, for now, that lesser-evil options exist. What would that mean for COST? COST states that one is required to choose the all-things-considered expectedly best act, unless it either involves unreasonable marginal expected costs to the agent relative to some alternative or else is better than every alternative only in virtue of the expected benefits to the agent.

Suppose you have two alternatives: turning the trolley towards one—the lesser evil—and letting the trolley kill five—the greater evil. By hypothesis, the lesser evil is expectedly morally better than the greater evil. So it is morally permissible to let the five die if and only if either (i) the lesser evil involves unreasonable marginal expected costs to the agent or (ii) the lesser evil is better than the greater evil only in virtue of a benefit to the agent that she is entitled to waive. On its face, pulling the lever to turn the trolley does not involve any significant personal cost or waiving of benefit, so it looks like COST should imply that inaction here is wrong.

We seem, then, to be forced to choose between COST and the existence of lesser-evil options. However, we can think of two other ways to vindicate the intuitions that otherwise support lesser-evil options, while retaining COST.

3.1 Lesser-evil options as agent-centred options

First, some will argue that lesser-evil options can be wholly reduced to agent-centred options: the decision to carry out the lesser evil (rather than let the greater evil happen) almost always involves significant personal cost; if it doesn't, then one does not have the option to act suboptimally. One could appeal, for example, to one's horror at having to perform the lesser-evil act, or to subsequent post-traumatic stress. We might also think that compliance with morality in such cases can be non-instrumentally costly. In the trolley case, if you pull the lever you become a killer. Isn't that a cost in itself?

In practice, perhaps some apparent lesser-evil options can be explained away like this. But we think it fails as an explanation of our core trolley cases. It will help to start by thinking about what kinds of personal cost can contraindicate moral requirements. Very broadly, we can understand these through one or other of the main candidate theories of well-being.[42] Roughly, these are various forms of hedonic or otherwise experiential well-being, preference-satisfaction theories (of many different stripes), and equally diverse objective list theories.

Hedonic or experiential theories of well-being are not a good pairing for COST. The miser is not plausibly subject to less exacting positive duties than the altruist, just because he laments every dollar he donates. To block a moral obligation, personal costs must be calibrated against some common, objective scale.

Individual trauma can perhaps be calibrated objectively, and different people might reasonably react differently to being put into a case like ours. But we can stipulate that away by designing the case so that it has no psychological effects at all. Whatever you choose, you won't see the results, and your mind will be instantaneously wiped of any recollection of the experience.

If the hedonic reading of COST doesn't help explain lesser-evil options, might an objectivist reading do better? We think not. In deciding which option is the lesser evil, we have already considered the objective reasons for and against pulling the lever and letting the five die. We have considered not only the lives that are at stake, but also the difference between intervening in the causal sequence, killing the one, versus letting it continue, so the five die. We have considered all this, and judged that letting the five die is objectively worse than killing the one. What, then, could explain the latter act being more objectively personally costly than the former? Every consideration that one might appeal to in reaching the latter judgement has already been appealed to in reaching the moral verdict—including the special importance to the agent of what she herself does.

On either an objectivist or an hedonic interpretation, then, COST lacks resources to explain why pulling the lever should be more costly than letting the trolley run. What about preference-satisfaction theories? We are sceptical. If killing the one is objectively better than letting the five die, it's hard to see why we should give any weight to a brute preference (i.e. one not grounded in any further considerations) that irrationally reverses that ordering. We seem to be back in Scrooge territory. It would be strange for two people to be faced with identical choices, but for only one of them to be required to pull the lever, just because the other has an objectively irrational brute preference. It's also strange to suppose that the moral importance of saving five lives cannot override your irrational preference to perform the worse act, when it *can* override the person on the side-track's very rational and weighty preference to *stay alive*.

What's more, as long as your preference for not turning versus turning the trolley is not lexically prior to all your other preferences, there must be some way to

[42] James Griffin, *Well-Being: Its Meaning, Measurement and Moral Importance* (New York: Oxford University Press, 1986); Derek Parfit, *Reasons and Persons* (Oxford: Clarendon Press, 1984): Appendix III.

adjust the case so that, if you turn the trolley, you receive a compensating benefit that makes you indifferent between turning (with the benefit) and not turning. In that event, you would be required to turn the trolley.[43] This too is counterintuitive.

In the end, we think that the reduction of lesser-evil options to agent-centred options fails because it places the wrong set of interests front and centre in your deliberations. Ultimately, what is at stake for you is very little in comparison with what is at stake for the six people tied to the tracks. Of course, deontologists like us think that you may rightly put more weight on your own actions than on what you let happen, and perhaps one might articulate a more expansive conception of the kinds of interests that ground lesser-evil options by appealing to these kinds of agent-relative reasons. But we have already taken that into account in the setup of the case. If you fail to act, five will die; if you act, only one will be killed. Taking *all* of your agent-relative reasons into account, as well as your agent-neutral reasons, it is all-things-considered objectively worse to let the trolley run its course than to turn it. So how then can it be worse *for you* to turn the trolley? To assimilate lesser-evil options to agent-centred options, we would have to argue that our moral judgements in such cases should be driven by your confessedly irrational preference for not turning over turning, rather than by the actual reasons at stake in your choice. This does not seem a palatable approach for deontologists. It is in the end strange to give so much weight in our deliberations to the irrational preferences of the lever-puller, relative to the much more serious, and clearly objective interests, of the people who will die if you do (or don't) turn the trolley.

3.2 Lesser-evil options and parity

Our agent-centred and lesser-evil options will, contingently, sometimes overlap. But we do not think the latter are plausibly reducible to the former. The second means of explaining away lesser-evil options while continuing to endorse COST is to argue that, when we think we have lesser-evil options, we are really just recognising that our alternatives are roughly equal to one another, or 'on par' (for our purposes, it does not matter whether we view this through the lens of parity, vagueness, or some other similar account). You are permitted to let the five die or kill the one, because neither option is clearly better than the other.[44]

Since we think incomparability of this kind is real, and has many interesting implications for both deontological decision theory in particular, and normative ethics in general, and since this argument has been made to us numerous times when we have presented this material to decision theorists, we think it is worth taking

[43] Note that while one difference between [not turn] and [turn plus benefit] is the additional benefit to you, which you are entitled to forgo, there is also the further difference that it is morally better to turn the trolley than not to turn it. So you would not have a self-sacrificing option to act suboptimally, according to COST.

[44] One could apply some of Caspar Hare's arguments about risky cases to reach this kind of conclusion, see e.g. Caspar Hare, 'Should We Wish Well to All?', *Philosophical Review* 125/4 (2016), 451–472.

seriously.[45] Ultimately, however, we think that it fails. And we think that showing this can illustrate one of the benefits of exploring deontological decision theory—it helps us to separate moral theories that are extensionally equivalent under full information.

It will help to begin with a trolley case with full information. This time six people are on the main track, one of them tied to the tracks 50 m in front of the others. There are two levers before you. If you pull the left-hand lever, then you will divert the trolley down a side-track, killing one person on that track, but saving the six who would otherwise have been killed. If you pull the right-hand lever, then the trolley will first kill the chap who is out on his own in front of the other five, and then divert to the same side-track, killing the one. Your options, then, are:

A. Do nothing, let the trolley kill six on the main track.
B. Divert the trolley before it hits the one, save all six prospective victims, but kill the person on the side-track.
C. Divert the trolley after it hits the one, killing him, saving the remaining five on the main track, and killing the person on the side-track.

If we have lesser-evil options, then A is morally permissible. And if you must choose between A, B and C, then C is clearly wrong. It is made wrong by the presence of B: if you're going to kill the one, then you must save as many lives as possible while doing so.[46] However, notice that if A and C were your only options, then both would be permissible. C would be the lesser evil. And since you have an option not to do the lesser evil, A would be permissible. Likewise, if A and B were your only options, both would be permissible.

Now, let's suppose that we want to represent apparent lesser-evil options by appealing to parity. We would then argue that, *in the choice between* [A, B, C], A and B are on par, while C is, in virtue of the presence of B, worse than both. For decision-making with full information, that could potentially work just fine. But let's now consider a risky version of the case.

Again, you can either do nothing or pull one of the two levers. But this time you don't know which lever is which. So your options are A, as above, and:

D. Pull the left-hand lever, which will with probability p do B, and with probability $1-p$ do C.
E. Pull the right-hand lever, which will do C if the left-hand lever would have done B, and B if the left-hand lever would have done C.

We have the following considered judgements about this case: if $p \geq 0.5$, option D is subjectively permissible. If $p \leq 0.5$, then E is permissible. A is subjectively

[45] Ruth Chang, *Incommensurability, Incomparability, and Practical Reason* (London: Harvard University Press, 1997).

[46] This is an instance of the 'Bang for Your Buck' principle, defended by Peter A. Graham in other work.

permissible for all values of p, but is never subjectively required. We think that the parity error theory for lesser-evil options cannot accommodate these judgements.

If parity explains the permissibility of doing A in the full-information version of the case, and if C is impermissible, then we have to rank the acts $A \approx B > C$ (where $\approx$ means 'is on par with'). Note that this is consistent with believing that, in the choice between A and C alone, $A \approx C$. In virtue of B's availability, C is now determinately worse than A, even if they are on par when compared in isolation from B.

The expected reason in favour of D is $pB + (1-p)C$. The expected reason in favour of E is the complement: $(1-p)B + pC$. Meanwhile the expected reason in favour of A is just A. If A and B are on par, but C is worse than both, then practically any probabilistic mix of B and C is going to be worse, in expectation, than A. The exception would be when the probability that pulling that lever actualises B is close enough to 1 that the expected reason in favour of pulling the lever is within the 'zone of parity' with A.

If this is right, then one is required to choose A, unless the risk that pulling a lever will actualise C is quite small. But this does not seem right. Inaction is subjectively morally permissible, for sure. But it does not seem morally required. You should be able to run *some* risk of acting suboptimally, for the sake of saving at least some of the six lives that will be lost if you do nothing. Indeed, as we suggested above, you should be able to choose D even if $p = 0.5$.

Perhaps one might reply here that C is objectively impermissible in the full-information version of the case just because *B is an alternative.* If your only options were A and C, then C would be permissible. One might then wonder whether the choice between A, D and E is really a probabilistic version of the choice between A, B and C, or else is a completely different choice, which cannot be understood in those terms. Of course, if you choose D, pull the left-hand lever, and it 'actualises' B, then objectively it is true that you had the alternative of doing C instead, because in that situation option E (pulling the right-hand lever) would have actualised C. But we might think that your objective moral reasons are sensitive to your information, in the sense that whether you act objectively wrongly depends on whether, if you end up doing C, the alternative of doing B was *epistemically* available to you. It *would* be seriously wrong to choose C when you knew that B was an alternative. But when your only options are D and E, then perhaps the seriousness of your objective wrongdoing, in the event that you pull the wrong lever, is mitigated by your not having a determinately better alternative. If this is right, then it should affect our judgement of which options are subjectively permissible.

We think there is some truth to this response—and that this offers still another pay-off of thinking through probabilistic cases like these, since they suggest how our uncertainty can sometimes affect our objective reasons. However, the objection still goes through. The fact that you lacked full information when choosing is relevant to the objective stakes. But choosing D (say) given that in fact it will actualise C is still objectively wrong, and we must assign that act–state pair less probability-weighted rational support than A, doing nothing, which does not risk objective wrongdoing.

One might object, instead, that it is wrong to write that $A \approx B > C$ in the full-information version of the case. Instead, we should write that $A \approx B$, and $A \approx C$, but

B > C. In other words, A is on par with both B and C, but C is impermissible because it is dominated by B (not because it is worse than A). In that case, as between D and E, it would of course be wrong to pull the lever with a lower probability of actualising B. But as between A and D or E, if B is on par with A, and C is on par with A, then a probabilistic mix of B and C should also be on par with A.

This response would require a departure from the decision rules that we have presented so far, and a further departure from classical decision theory.[47] We cannot, here, consider the merits of that whole approach. So here is a further argument, aimed at those who are inclined to choose this second route to reach the subjective permissibility of doing nothing in the risky version of the first case. This argument also involves duties that are not as obviously sensitive to one's epistemic position as one's duties of beneficence.

In general, we tend to think that imposing costs as a lesser evil is permissible only if the moral benefit realised is proportionate to the cost imposed (only then would it genuinely be a lesser evil). While it may be permissible to turn the trolley towards one to save five, it is probably impermissible to turn the trolley towards two to save five.[48] In general, we think that one's duties not to harm others are not as sensitive to one's epistemic position as one's duties of beneficence. If your action causes harm to another person, to which they are not liable, then you have breached a duty to them even if there was no way you could have known that your action would have that result.[49] Suppose now that you face this choice:

A. Do nothing. Five die.
B. Divert trolley down side-track B, killing one and saving five.
C. Divert trolley down side-track C, killing two and saving five.

By hypothesis, we have lesser-evil options in cases where our only alternatives are A and B. The availability of C doesn't change that. So A and B are both objectively permissible. C is objectively impermissible. Notice that C would be objectively impermissible even in a choice between only A and C. A and C are explicitly *not* on par.

Now consider the risky version of this case, where your options are A, or

D. Pull the left-hand lever, which will with probability p do B, and with probability $1-p$ do C.
E. Pull the right-hand lever, which will do C if the left-hand lever would have done B, and B if the left-hand lever would have done C.

[47] Caspar Hare, 'Take the Sugar', *Analysis* 70/2 (2010), 237–247.

[48] In fact, we disagree on how weighty the doing/allowing distinction is here. We have chosen the ratio of harm inflicted to harm averted that deontologists typically consider proportionate, but the case would work just as well if C involved killing up to four people.

[49] Thomson, *Rights, Restitution.*

By hypothesis A $\approx$ B > C, so either D or E, which are probabilistic mixes of B and C, are guaranteed to be expectedly worse than A, provided the probability of C is above the low threshold that preserves the rough equality. Unless the penumbral zone is quite wide, this means that we can tolerate practically no additional risk of killing an innocent person in order to save five lives.

We think that this result alone is implausible. It should be possible to run some risk of killing an additional innocent person in order to save five lives. But even if you disagree, there is a further problem. If we represent apparent lesser-evil options in this way, then our decision in risky choices like this must remain the same, even if we increase the number of lives at stake, provided A and B continue to be on par. In standard trolley cases, many would agree that it is merely permissible to divert the trolley towards one, even if you could save 20 by doing so. Imagine, then, that you have the following options:

A. Do nothing: 20 die for sure.
B. Divert trolley down side-track B, killing one and saving 20.
C. Divert trolley down side-track C, killing two and saving five, letting 15 die.[50]

C is clearly worse than B, which dominates it. And it is worse than A—the difference between them is the same as in the former case, and it is not permissible to kill two as a side-effect of saving five. So, as above A $\approx$ B > C. Now suppose that you can choose between A and

D. Pull the left-hand lever, which will with probability p do B, and with probability $1-p$ do C.
E. Pull the right-hand lever, which will do C if the left-hand lever would have done B, and B if the left-hand lever would have done C.

We think that, in this choice, you should be permitted to run *more* of a risk of actualising C than was permissible in the previous case, when only five lives were at risk. The prospect of potentially saving 20 lives should justify running a *greater* risk of objective wrongdoing. But in fact, the parity-based approach delivers the opposite verdict. A and B must remain on par, to adequately represent the lesser-evil option to do nothing in the full-information version of the case. But C is presumably worse than B by a greater degree than when only five lives are at stake (as well as killing one additional person, you're failing to save an additional 15). So options D and E are not just wrong, they are more seriously wrong than they were before. This is exactly the wrong result.

[50] The arrangement of the tracks would have to be a little complicated for this case to work. The 20 victims are in two groups. 15 are at the end of the main track, 5 are ahead of them. If you do nothing, the train will kill all 20—first running over the 5, then the 15. If you pull the lever, you can send the trolley down either B or C. If it goes down B, then it kills 1 and comes to a halt. If it goes down C, then it kills 2 and continues on a loop back onto the main track, rejoining after the group of 5, and running on to kill the 15. So you save the 5 at the expense of the 2, while still letting the 15 die.

Might one respond, here: so much the worse for the commonsense view of lesser-evil options? Perhaps our opponent could simply argue for a narrower zone of parity, and deny that B and A are on par. But we think the objection is still compelling even in light of that modification. If we change the case around so that the best scenario is that you save six lives, we still think you should be entitled to run a greater risk of the worst-case outcome coming about to save six than you would be allowed to do in order to have a chance of saving five. One additional life at stake *should* make a difference to your decision. Its being swallowed up in the rough equality is not only counterintuitive, it reveals an objectionable flaw in the underlying theory. It is simply not plausible that options that differ to such a marked degree in terms of what matters can defensibly be represented as being on par.

The standard cases used to spark intuitions about parity involve adding a trivial saving to the decision to have either Chinese or Indian food tonight. A few dollars saved is a ‘sweetener’, but in our cases we’re changing the case by adding *an extra life* that you can save. An extra life is a very big deal, not the kind of ‘sweetener’ that can plausibly be used to indicate parity.

More generally, if we want our moral decision theory to be generative, and not merely to represent judgements of which we are already confident, then we should aim to represent our moral reasons as faithfully as we can—while still rendering them amenable to combination with probabilities. All of these cases suggest that we should represent lesser-evil options in such a way that we respect two basic points: inflicting a lesser evil is better (that is, more supported overall by one’s reasons) than doing nothing; yet it is merely permissible, not required. The parity error theory cannot do this. It is a technical fix for a substantive moral problem.

4 Justifying lesser-evil options

If we have lesser-evil options, then existing versions of moral decision theory cannot adequately accommodate them. So: do we have lesser-evil options? We cannot offer a conclusive answer here, but we can indicate a direction of travel.

One can vindicate lesser-evil options in more or less extreme fashion. First, one could argue that the very *notion* of moral requirement is mistaken. We still have moral reasons, of course, and actions can be morally better or worse. But our moral reasons cannot make an action wrong—or for that matter required. If this is right, then lesser-evil options follow trivially: one is merely permitted, not required, to perform the lesser evil, because one is not *required* to do anything. Scalar consequentialists adopt this view.[51] In principle, deontologists could endorse it also. But they are unlikely to do so—wrongdoing is at the heart of deontological ethics.

One could take a less extreme approach, and argue that while *some* reasons can only count in favour of an action, others can make an action required, or that

[51] Gerald Lang, ‘Should Utilitarianism Be Scalar?’, *Utilitas* 25/1 (2013), 80–95.

individual reasons can have different levels of requiring or justifying force.[52] Perhaps, for example, our reasons to aid others can make one act morally better than another, but not make it wrong to perform the worse act. And maybe our reasons not to harm others can both favour and require, so that not harming someone is morally better than harming them, *and* the latter act is wrong.[53] Again, if this were right, then we could defend at least some lesser-evil options. In the original trolley case, for example, perhaps your justifying reasons to save the five override your requiring reason not to harm the one, so that it is permissible to turn the trolley, but not required.

Of course, if this approach were right, then no matter how many lives you could save, you would not be required to turn the trolley. Some deontologists will be comfortable with this result, but we want to vindicate something more in tune with commonsense morality. It is possible, of course, that any given fact in some context can have some measure of justifying weight, and some measure of requiring weight. So we wouldn't need a crude hard-and-fast rule that 'aiding only favours, while not-harming can require'. For present purposes, though, we want to identify an argument that more immediately vindicates lesser-evil options, rather than zooming out to 36,000 ft.

We can see a path to such an argument. It proceeds via two claims. First, we think that to ground requirements we must meet a higher justificatory burden than to ground permissions. Second, we suggest that the first claim is in part explained by facts about moral endorsement and condemnation. If the first claim is right, then we have an intuitive case for lesser-evil options. If the second claim is right, then that intuitive case has robust foundations.

Here, then, is the minimum viable argument. We think that, in general and other things equal, it is harder to justify a requirement to ϕ than a permission to ϕ. For example, the justificatory burden that one must meet for it to be permissible to harm one person for the greater good is less than must be met for it to be required to do so. That turning the trolley will save five lives is enough to make killing the person on the side-track permissible, but not enough to make it required.

Consider a target action, ϕ. Our claim is that, other things equal, it is harder for the reasons in favour of ϕ to establish that ϕ is required than it is for them to establish that ϕ is merely permissible. For the former to be the case, the reasons for ϕ must be able not only to ground that ϕ is permissible, but also to show that the alternatives to ϕ are impermissible. For the latter to be the case, the reasons for ϕ need only establish that ϕ is permissible.

[52] There are many different attempts to characterise this kind of distinction in both moral philosophy and practical reason. The most influential is Joshua Gert's: see his 'Requiring and Justifying: Two Dimensions of Normative Strength', *Erkenntnis* 59/1 (2003), 5–36; 'Normative Strength and the Balance of Reasons', *The Philosophical Review* 116/4 (2007), 533–562; 'Practical Rationality, Morality, and Purely Justificatory Reasons', *American Philosophical Quarterly* 37/3 (2000), 227–243.

[53] Robert Nozick, *Anarchy, State and Utopia* (Oxford: Basil Blackwell, 1974); Warren S. Quinn, 'Actions, Intentions, and Consequences: The Doctrine of Double Effect', *Philosophy and Public Affairs* 18/4 (1989), 334–351.

Now, of course it is possible for ϕ to be required not because the reasons in favour of it are so strong, but because there are very strong reasons against all the alternatives. And sometimes the reasons in favour of a permissible act are very weighty indeed—for example, in some cases of supererogatory action. Our core claim is simply that, *other things equal*, it is harder to justify a requirement than a mere permission.

If you find that idea appealing, then we're already most of the way to vindicating lesser-evil options—we have them just because, in general, it is harder to ground a requirement than to ground a permission, and lesser-evil options arise when the difference between the two options is great enough to justify a permission to perform the lesser evil, but not great enough to make it required. But it would help to explain *why* requirements are harder to justify than permissions, other things equal.

We can offer some preliminary considerations to that effect, tailored to the case of lesser-evil options. We think that for one to be *required* to inflict a given cost C on someone else signals a greater degree of moral endorsement of C than if one is merely *permitted* to inflict C. It is harder to justify *that* additional degree of moral endorsement. We think that this ties the existence of lesser-evil options to the general kinds of rationale that explain why, when counting whether a given harm counts as a lesser evil at all, we do not simply weigh the harms against one another, but instead pay attention to facts like whether the harm in question is intended or merely foreseen, and if foreseen how likely it was to occur, and whether it is or isn't a causal means to removing the greater evil, and so on.

We think that all of these considerations are grounded, very roughly, in the fact that beings with moral status are ends in themselves, and enjoy a certain kind of inviolability, which protects them against bearing costs that they are not liable to bear, for the sake of the greater good.[54]

Suppose that person S_1 can intervene to prevent S_2 from suffering a given cost C. Doing so, however, would be personally costly. It is quite consistent with S_1 being morally permitted to let S_2 suffer C that C is wholly morally condemned. It may be entirely morally objectionable that S_2 suffer C, and yet given the cost to S_1 of intervening, she is permitted to let C occur. This is, we think, because while S_2 may suffer a violation if C occurs, that violation does not compromise his status as inviolable. To be inviolable is *to be such that you ought not be violated*. S_2 is still inviolable.

Now suppose that S_1 inflicting a given cost C on S_2 is unavoidable if she is to realise a benefit B for S_3. If it is permissible for S_1 to inflict C on S_2 for the sake of realising B for S_3, then S_2 *is not* such that he ought not be violated. *This* violation, or infringement, *is* permissible. It cannot therefore be represented as wholly morally condemned. It might be regrettable, or pro tanto wrong, but it is at least somewhat morally endorsed. Note, though, that since C is unintended, and so not a causal means to realising B, that implies that C is somewhat *less* endorsed than would be true if those facts did not hold. If C were an intended means to B, then if it

[54] Warren S. Quinn, 'Actions, Intentions, and Consequences: The Doctrine of Doing and Allowing', *Philosophical Review* 98/3 (1989), 287–312; F. M. Kamm, 'Review: Non-Consequentialism'.

is permissible to inflict C for the sake of B, that signals a still greater endorsement of S_2 suffering C. S_1 is actively permitted to develop a plan that involves using S_2 and his suffering as a means to the greater good. This is a more serious incursion into S_2's inviolability than when the harm is an unintended side-effect.

We contend that permissions and requirements work in the same kind of way, in parallel with considerations of omission and action, intention and causal connection. Other things equal, if it is merely permissible for S_1 to inflict C on S_2 for the sake of realising B for S_3, that signals a lesser degree of moral endorsement of C than if S_1 is required to inflict C on S_2 to realise B. S_2's inviolability is somewhat less compromised when it is merely permissible to harm him to help S_3; it is more compromised when harming S_2 is not only permissible, it is required. He is 'there for the sake of others' to a greater degree if *not* harming him is not an option.

Prima facie, these considerations seem relatively discrete and additive, at least when we hold C constant. If S_1 is morally required to intentionally cause C to S_2 as a means to realising B for S_3, then that constitutes a more or less maximal moral endorsement of C, and so of using S_2 for the benefit of others. For this to be morally required, B would have to be significantly greater than C. Conversely, if S_1 is merely morally permitted to unintentionally let C happen to S_2, because it is a side-effect of realising B for S_3, then that need not constitute any moral endorsement of C at all, so S_1 is plausibly allowed to simply weigh C against B, and bring about B if it is no less weighty than C. We will not attempt to parse how these considerations interact with one another in more complicated cases. We also note that if S_2 is liable to bear C, that might generate complex interaction effects—we cannot explore those here.

For S_1 to be required to impose C on S_2, when S_2 is not liable to bear C, implies that C is more endorsed than if imposing C is merely permissible, other things equal. We have two additional arguments for this claim. First, requirement entails permission, but not vice versa.[55] Other things equal, any degree of moral endorsement that attaches to mere permission should therefore attach also to requirement. The requirement to impose C involves something beyond a mere permission: one is not permitted *not to* impose C. This extra element implies additional endorsement; it elevates imposing that cost above its alternatives. To say that it is optional is to say: you may do that, and you may do some other thing also. The optional act is not thereby elevated above the alternatives.

Second, we think that the normative upshots of being required to impose C are different from those of being merely permitted to do so. If S_1 is merely permitted to impose C on S_2, for the sake of realising some moral benefit B, then S_1 may be to a greater extent liable to compensate S_2 for C than would be true had she been required to impose C.[56] If imposing C is merely permissible, then S_1 has a reasonable alternative to imposing C. If imposing C is morally required, then S_1 does not have a reasonable alternative, since all alternatives involve wrongdoing. In general, if one has a reasonable alternative to performing some action, and one nonetheless performs the action, then other things equal one is more responsible for bearing the costs, should

[55] In terms of deontic logic, requirement implies truth in more possible worlds than permission, therefore it is harder to come by. Thanks to Alan Hájek here.

[56] Thanks to Christian Barry for helpful discussion here.

that action eventuate in harm, than if one had no reasonable alternative to performing that action.[57] Of course, in most cases of lesser-evil harming, other beneficiaries may also bear the cost. And we might also want to incentivise aiming at the greater good by sharing the costs of compensation more broadly around society. We mean only to suggest that, other things equal, the fact that one's compensatory obligations are less after imposing a required cost than after imposing a merely permissible cost lends support to the idea that the former costs are more morally endorsed than the latter.

Similarly, we think it is often permissible to intervene to prevent someone imposing a cost when they are merely permitted to do so, when it would be impermissible to intervene if they were morally required to bring about the same cost.[58] In passing, this might shed some light on a debate in the ethics of war, over whether innocent civilians whose lives are threatened by necessary and proportionate enemy bombing are permitted to defend themselves, even though it would prevent their attackers from fighting proportionately for a just cause. We suggest that if the attackers are morally required to carry out those bombing raids, then it will be harder to justify using lethal defensive force against them, than if they are merely permitted to do so.[59] This, too, lends support to the idea that required costs are morally endorsed in a way that merely permitted costs are not, since one may prevent the imposition of the latter under conditions in which one would not be permitted to prevent the imposition of the former.[60]

The argument for lesser-evil options, then, is this: beings with moral status are ends in themselves, who enjoy a certain inviolability against bearing costs for the sake of the greater good. Some incursions into their inviolability are nevertheless justified, but more severe incursions are harder to justify than less severe incursions, other things equal. It is obviously harder to justify imposing a greater than a lesser cost on an inviolable person (when other things are equal, and in particular when that person is not liable to bear that harm). But it is also harder to justify imposing costs that imply a greater degree of moral endorsement for the use of that person for the sake of the greater good. More goes into determining the degree of moral endorsement than just the magnitude of the cost. Other things equal, if it is permissible to intend harming a person, that amounts to a greater moral endorsement of the harm than if it is permissible only to harm them unintentionally; if it is permissible to do harm to a person, that is a greater endorsement of that harm than if it is merely allowed; and if one is *required* to harm a person, that amounts to a greater endorsement of that harm than if it is merely permitted. Greater endorsements are greater

[57] Stephen Perry, 'The Moral Foundations of Tort Law', *Iowa Law Review* 77 (1991–1992), 449–514; H. L. A. Hart and Tony Honoré, *Causation in the Law* (Oxford: Clarendon Press, 1985).

[58] We owe this idea to conversations with Lars Christie.

[59] For discussion, see Jeff McMahan, 'Debate: Justification and Liability in War', *Journal of Political Philosophy* 16/2 (2008), 227–244.

[60] We note that this suggests a further argument against the attempt, above, to reduce lesser-evil options to agent-centred options. Suppose S_2 can prevent S_1 from harming him, and so deprive S_{3-n} of the benefit they would otherwise have had through S_1's action. We suggest that if S_1 is required to impose that cost on S_2 for the sake of S_{3-n}, then it is harder for S_2 to justify preventing S_1 from acting than if S_1 is merely permitted to impose that cost on S_2. The fact that S_2 is permitted to resist S_1, in some case, but required not to do so in others, seems to have everything to do with the interests of S_2 and S_{3-n}, and very little to do with S_1's interests.

incursions into the victim's inviolability, and as such are harder to justify. We have lesser-evil options in cases where the good achieved is great enough to justify a permission to harm the victim, but not great enough to ground a requirement to do so.[61]

5 Accommodating Lesser-evil Options

Deontological decision theorists were originally attracted to classical decision theory because of both its simplicity and its explanatory power in other domains. But its commitment to ordering act–state pairs along a single dimension is its downfall. This is clear once we take our moral licence to favour or thwart our own interests adequately into account.

We must at least find ways of ranking act–state pairs with respect to both moral betterness, and personal cost—and figure out what degree of moral improvement can make a given degree of personal cost morally required. But introducing this further dimension does not help accommodate lesser-evil options, nor can they either be reduced to agent-centred options or else explained away as artefacts generated by subtle incomparabilities. This leaves us with three possible courses of action.

First, we could give up on deontological decision theory entirely. This might involve endorsing a belief- or knowledge-first approach to moral decision-making under uncertainty.[62] Alternatively, we could try to make a domain-specific approach work.[63] We think these alternatives are certainly worth exploring. But there are legs in deontological decision theory yet.

[61] Despite appearances, our view of lesser-evil options is ultimately quite similar to Helen Frowe's 'Preventing Harm' principle: 'One has a duty to prevent harm to others when one can do so without violating anyone's rights, and without bearing an unreasonable cost.' Our principle is more general than Preventing Harm. But, in Frowe's terms, our view is roughly that one has a duty to prevent harm to others when one can do so without imposing unreasonable costs on oneself or on others. In other words, where she focuses on violations of rights, we focus on the imposition of unreasonable costs. This makes our view more flexible, and allows it to accommodate the intuitively correct verdict on the following case (which Frowe's view cannot accommodate):

> A trolley is headed toward one person, Bill. It can be diverted onto a side-track on which there is another person, Bob. Bob has freely and uncoercedly consented to having the trolley turned towards him to save Bill. (It's not that Bob wants the trolley turned on him—he'd prefer that it not be turned on him, actually; rather, he merely consents to its being turned on him.) A bystander can divert the trolley or not.

We think it is quite likely that it is permissible, but not required, for the bystander to turn the trolley in this case. So she does not have a duty to prevent the harm to Bill even though she could do so without violating anyone's rights (Bob has waived his right not to be killed and so turning the trolley on him would neither violate, nor even infringe, a right not to be killed) and without bearing an unreasonable cost. She would, however, be imposing an unreasonable cost on Bob—albeit one to which he has consented. So our approach would cater for this case.

[62] Tenenbaum, 'Action, Deontology, and Risk'; Isaacs, 'Duty'.

[63] E.g. Jonathan Quong, 'Rights against Harm', *Aristotelian Society Supplementary Volume* 89/1 (2015), 249–266; Tomlin, 'Subjective Proportionality'; Renée Jorgensen Bolinger, 'Reasonable Mistakes and Regulative Norms: Racial Bias in Defensive Harm', *Journal of Political Philosophy* 25/2, 196–217.

Second, we could introduce yet another dimension into our deontological decision theory, and hope that it doesn't make things too complicated (and that there are no further dimensions to take into account). And third, we can try to identify the underlying structure of both agent-centred options and lesser-evil options, and encompass them both within a simpler single principle. We will attempt the second approach in this section; the third would take a paper in its own right, though we think it may ultimately prove necessary.

Here is a principle that accommodates lesser-evil options organically, and can deal with all the problem cases raised above.

COST+: An act is subjectively permissible for an agent if and only if:

(a) there is no all-things-considered expectedly better act or
(b) every all-things-considered expectedly better act

    (i) involves unreasonable marginal expected costs to the agent, or
    (ii) is better only in virtue of expected benefits to the agent, or
    (iii) involves imposing unreasonable marginal expected costs on some others.

COST+ adds a further clause to COST, noting that an act might be permissible because the alternatives involve imposing costs on others that one is not required to impose, despite the additional moral benefit that they can yield. Its key difference from COST is that, where COST ranks options by considering both their overall support by moral reasons and their impact on the agent, COST+ also factors in costs that are imposed on others.

We could make COST+ even simpler by merging (iii) and (i) into a single condition. We are not sure about this move, however, both because we think it is important to allow room for what counts as 'unreasonable' being different when the costs are to the agent and to others, and because we are not sure that in every case where one has an agent-centred option to avoid some cost, one strictly speaking has an option to avoid *imposing* that cost on oneself. Consider, for example, the difference between being required to turn the trolley on oneself, and being required not to harmlessly prevent the bystander from turning the trolley towards oneself.

COST+ allows the possibility that inflicting harm as a lesser evil can genuinely be better than doing nothing, without implying that one is required to inflict the lesser-evil harm. Consider the first case that proved problematic for COST:

A. Do nothing: let the trolley kill six on the main track.
B. Divert the trolley before it hits the one, save all six prospective victims, but kill the person on the side-track.
C. Divert the trolley after it kills the one, saving the remaining five on the main track, and killing the person on the side-track.

We can now rank these options as the reasons at stake suggest that we should: both B and C are better than A. C, however, is impermissible not simply because it is outranked by B, but because B involves imposing no additional cost (and involves no additional cost to the agent) and is morally better. A is permissible because both B and C involve imposing unreasonable marginal expected costs on the one bearing the cost.

Now suppose that we can choose between A and

D. Pull the left-hand lever, which will with probability p do B, and with probability $1-p$ do C.
E. Pull the right-hand lever, which will do C if the left-hand lever would have done B, and B if the left-hand lever would have done C.

For any value of p, A will be subjectively permissible. Any probabilistic combination of B and C involves imposing unreasonable expected cost relative to A. If $p=0.5$, then both D and E are permissible. If $p>0.5$, then D is permissible, and E is impermissible, because it is outranked by an option (D) that does not involve imposing additional expected cost. If $p<0.5$, then E is permissible, D impermissible because it is outranked by an option (E) that involves imposing no more expected cost.

Return next to the case where there was some risk of acting clearly impermissibly, but that has to be weighed against the good that could be done if you inflict the lesser-evil harm.

A. Do nothing: let 20 die.
B. Divert trolley down side-track B, killing one and saving 20.
C. Divert trolley down side-track C, killing two and saving five, letting 15 die.[64]

Here B > A > C, as it should be. Nonetheless both A and B are permissible, because, though B is better than A, it involves imposing unreasonable marginal expected cost. Now suppose your options are A and

D. Pull the left-hand lever, which will with probability p do B, and with probability $1-p$ do C.
E. Pull the right-hand lever, which will do C if the left-hand lever would have done B, and B if the left-hand lever would have done C.

Now, when considering whether D is subjectively permissible, we need to weigh the prospect of saving an additional 15 lives against the risk of killing one additional person. We do not know for what values of p one is permitted to divert the trolley. But that value will be determined by weighing the right considerations: all the lives at stake will count, as will all the potentially imposed costs.

[64] See footnote 51 for how this would work.

These cases raise a further interesting question for COST+: namely, what should we make of cases in which the person bearing the imposed cost is different from one scenario to the next. For example, suppose your options are

A. Do nothing: let 20 die.
B. Divert trolley down side-track B, killing one and saving 20.
C. Divert trolley down side-track C, killing *a different* two and saving five, letting 15 die.

When thinking about agent-centred options, this question did not arise, because obviously the person bearing the cost in the different scenarios was the agent. But how should we think about the imposed costs in these cases? The intuitively plausible approach is to think about this, again, from the agent's perspective: what matters is the amount of cost that she is imposing on others. But this doesn't work so well, given the rationale for lesser-evil options. Suppose one option is expectedly better than another. In the simplest case, if the person bearing the cost is the same in both scenarios, we can ask whether *that person* is required to bear the difference in cost between the two scenarios for the sake of the additional moral benefit. That's the ideal scenario, but obviously we might also face a case in which the person bearing the imposed cost in the other scenario is someone different. We might also generate the same expected cost by having more or less widely dispersed risks.

On the first point, we think that what matters here may be a kind of impartiality. From the agent's perspective, given that everything else is equal, it doesn't matter on *whom* she imposes the costs. We can therefore treat the person bearing the cost as a kind of anonymous placeholder.[65] The question is not whether *this individual* is required to bear the cost, but whether the person bearing the cost (whoever that may be) is required to do so. The person bearing the cost plays a similar role in COST+ to the 'worst-off person' in Rawls's difference principle.[66]

We can reflect the second point either in how we calculate the interests of those exposed to these risks, or in their moral weight. At least one of us thinks there is a weak pro tanto reason to favour more widely dispersed risks. COST+ can accommodate that point.

6 Conclusion

Deontologists are only now awakening to the problem of imperfect information. The decision-theoretic route is not the only one available to them. And, as we have shown here, they had best not adopt classical decision theory wholesale if they want to preserve central deontological commitments, in particular to a range of options

[65] We are both uneasy with this possibility, and think there may be more to be said here; for reasons of space, however, we leave further discussion for a different occasion.

[66] Of course, if the different options *do* involve the same person bearing costs, that makes a difference—other things equal we have reason to disperse rather than concentrate costs.

to act suboptimally—beyond those grounded in their legitimate authority over their own interests. The partnership between deontology and decision theory must be just that—a partnership, as distinct from a takeover. We already knew that single-ranking approaches to decision theory were inadequate to accommodate deontological ethics; we've shown here that dual-ranking alternatives are also problematic. If we have lesser-evil options, then we at least need a third ranking, of imposed cost, which stands alongside personal cost and overall moral reason. Of course. one could deny that there are lesser-evil options, but we think they are an appealing feature of commonsense deontological morality, and that it makes sense that it should be harder to justify requiring the imposition of a cost, than merely permitting it.

Are we fully confident, however, that we have exhausted the variety of possible options to act suboptimally? Are we certain that there are no more dimensions along which act–state pairs may be ranked? We are not, and we doubt whether it is sustainable to keep adding in subclauses to accommodate each exception to the general rule of maximisation. We conjecture (and one of us argues elsewhere) that the solution may be to move up a level of abstraction, to identify the dimension of normative strength that unifies both agent-centred options and lesser-evil options, without implying that one is reducible to the other.[67] Of course, the cost of making our principle more abstract is that any gains in extensional adequacy are offset by losses in explanatory power, as well as in action guidance. Even if COST+ gets some cases wrong, it may prove more illuminating than a more abstract counterpart, or one with more epicycles.

Even if we have identified all the relevant options to act suboptimally, COST+ may need further refinement. After all, it purports to provide necessary and sufficient conditions for an act being subjectively permissible: that's a big aspiration, and not one that can feasibly be vindicated in a single paper. However, it does provide a proof of concept, to show that deontologists who endorse lesser-evil options, but want to develop a robust decision theory, have at least one promising avenue to pursue.

Funding Funding was provided by Australian Research Council (Grant No. Dp170101394).

[67] Here is the principle that Seth Lazar plans to defend elsewhere: An act ϕ is subjectively permissible just in case there is no alternative ψ such that your probability-weighted duty to ψ rather than ϕ outweighs your probability-weighted permission to ϕ rather than ψ (in *Duty Under Doubt*, a monograph project with Oxford University Press).