

HOW TO MODEL LEXICAL PRIORITY

Martin Smith

University of Edinburgh

A moral requirement R_1 is said to be *lexically prior* to a moral requirement R_2 just in case we are morally obliged to uphold R_1 at the expense of R_2 – no matter how many times R_2 must be violated thereby. While lexical priority is a feature of many ethical theories, and arguably a part of common sense morality, attempts to model it within the framework of decision theory have led to a series of problems – a fact which is sometimes spun as a ‘decision theoretic critique’ of lexical priority. In this paper, I develop an enriched decision theoretic framework that is capable of modelling lexical priority while avoiding all extant problems. This will involve introducing several new ingredients into the standard decision theoretic framework, including multidimensional utilities, de minimis risks, and the means to represent two different conceptions of risk.

Keywords: lexical priority, expected utility, multidimensional utility, de minimis risk, probabilistic risk, normic risk

1 Lives Over Headaches

If offered a choice between taking the life of an innocent person and inflicting a short-lived, mild headache upon each of a number of people, some would say that we are always morally required to choose the latter option, no matter how many people are involved. On this view, no number of headaches could ever outweigh or counterbalance the value of a life – preserving life takes *lexical priority* over the avoidance of headaches. Principles like this concern decisions under *certainty* – they tell us what morality requires of us in a situation in which the outcomes of our choices are stipulated to be certain. According to some critics, however, the idea of lexical priority breaks down when it comes to decisions under *risk* – in which we have imperfect information about how our choices will turn out. What if avoiding the headaches incurs only a 50% chance that an innocent will die? Am I still morally required to inflict the headaches, no matter the number? What if the chance is only 20%, or 1% or 0.000001% ...? The most widely accepted formal framework for navigating decisions under risk is that provided by *decision theory*. According to some critics, any attempt to accommodate lexical priority within this framework will force us to stomach a series of unpalatable results, including highly implausible answers to the above questions (Jackson and Smith 2006, 2016; Huemer 2010; Colyvan, Cox and Steele 2010: section 3; Hansson 2013: chap. 2). This is sometimes referred to as the ‘decision theoretic critique’ of lexical priority (Lee-Stronach 2018).

My aim here is not to argue that some things take lexical priority over others. My aim is not even to defend the possibility of lexical priority from the decision theoretic critique – though it will turn out that the critique, as it stands, trades upon modelling assumptions that are by no means obligatory. In this paper I will use the idea of lexical priority as a pretext for exploring several ways in which decision theory can be expanded and enriched. My primary aim is to develop and defend a new

decision theoretic framework which incorporates elements such as multidimensional utilities, de minimis risks and the means to represent two different conceptions of risk. The plan is as follows: In the next section I will explain the decision theoretic critique of lexical priority. More precisely I will outline two problems that are thought to beset any attempt to capture lexical priority within decision theory – the problem of *permissiveness* and the problem of *risk*. In the third section I will outline a strategy for solving the problem of permissiveness, linked to the idea of multidimensional utilities – utilities that capture several distinct and irreducible axes of value. In the fourth section I will outline a strategy for solving the problem of risk, linked to the idea of a de minimis risk – a risk that is so small it may be legitimately ignored for the purposes of decision making. In the fifth section, I discuss a problem associated with the introduction of de minimis risks into decision theory – one which could, at worst, precipitate the collapse of the entire framework.

Up to this point, my paper will follow the contours of Chad Lee-Stronach's excellent 'Moral priorities under risk' (2018). In particular, the idea of introducing multidimensional utilities to solve the problem of permissiveness, and the idea of introducing de minimis risks to solve the problem of risk are both pioneered by Lee-Stronach. Lee-Stronach is also aware that the second strategy has the potential to de-stabilise the decision-theoretic framework – though my way of shoring up the framework is rather different from his. In the sixth section, I take a new turn, outlining a novel conception of risk that I term the *normic account*. On the normic account, the risk of a given outcome is determined by its *abnormality* rather than its *probability*. In the final section, I argue that one way to successfully model lexical priority, while safeguarding the stability of the framework, is to equip decision theory with the means to represent both normic and probabilistic risk, and to define de minimis risks in normic terms. The appendix provides some supporting technical results.

2 Lexical Priority and Expected Utility Theory

Lexical priority is most naturally thought of as a relation between moral requirements or imperatives. We might say that a moral requirement R_1 is lexically prior to a moral requirement R_2 just in case a single violation of R_1 is reckoned morally worse than any number of violations of R_2 . If R_2 is the requirement that one not inflict mild headaches and R_1 is the requirement that one not take the life of an innocent person, then we arrive at the 'lives over headaches' principle from the last section. While this suffices as a basic definition, for what follows we will need to rework it somewhat, in order to better fit with the vocabulary of decision theory.

Lexical priority claims, of varying plausibility, are a feature of many ethical theories. For instance: Kant appeared to hold that one is never permitted to lie, even to protect others from harm (Kant 1797; Korsgaard 1986). Starkie claimed that it is always impermissible to deliberately convict an innocent person of a crime no matter how many guilty people one might convict in the process (Starkie 1824: pp572-574). Dworkin held that one must never infringe an individual's rights in order to pursue

general societal benefits or avoid societal costs (Dworkin 1977: chap 4, section 3, chap 7).¹ The ‘lives over headaches’ principle, which is often used as a stand-in for lexical priority claims in general, is described as ‘obviously true’ by Thomson (1990: p169) and is sometimes said to be a part of common sense morality (see, for instance, Brennan 2006: p251; Dorsey 2009: pp36-37; Dougherty 2013: p1; Kirkpatrick 2018: p107). The principle is not universally accepted, though, and it’s prominence in the literature may be partly due to Norcross’s (1997, 1998) attempts to argue against it.

The decision theoretic frameworks I consider here will use the following elements to model a decision problem: First, we require a pairwise exclusive and jointly exhaustive set of available *actions*. Second, for each action we require a pairwise exclusive and jointly exhaustive set of possible *outcomes* that could result. Third, we need a *risk* function which assigns some measure of risk to each outcome, given each action. Fourth, we need a *value* function which assigns some measure of value to each outcome. The final ingredient is a *decision rule* which selects an action or actions, given the outputs of the risk and value functions.²

Decision theory is sometimes given a subjective interpretation on which the value function and risk function are supposed to represent the views of a particular agent (who conforms to certain very broad constraints of rationality). On this interpretation, the decision rule selects the action or actions that are rational for that agent, given their own preferences and risk estimates. Our interest here, however, is with an interpretation on which decision theory offers *moral guidance*. For this purpose, we should take the value function to reflect objective values and the risk function to reflect the risk estimates that are supported by an agent’s evidence. On this interpretation, the action or actions that are selected by the decision rule are those that the agent is morally permitted to choose, given their evidence.³ I don’t assume here that the decision rule is something that an agent could necessarily *follow* when making decisions – at least not in all cases. It’s enough that the rule provide a means for us, as theorists, to identify the morally permissible options.

¹ Advocates and critics of lexical priority views don’t always use the *term* ‘lexical priority’. The term was coined by Rawls to describe one possible response to the problem of reconciling competing principles of justice (Rawls, 1971: chap. 1, section 8).

² A model of a decision problem will sometimes include, in addition, a set of *states* of the world, which are relevant to determining the outcomes of the available actions and which lie beyond the agent’s control. While states won’t feature in the formalism used in the main text, I return to them in n6, n16 and n26.

³ One might think that there is a fundamental mismatch between lexical priority and decision theory, on the grounds that the latter presupposes a *consequentialist* approach to morality, while the former is most at home within a non-consequentialist framework. One commitment that we do need to undertake, when modelling lexical priority in decision theoretic terms, is that lexical priority relations are grounded in underlying *value differences*: If we are morally obliged to inflict any number of headaches rather than a death, this is because a death has *in some sense* a greater disvalue than any number of headaches. Some might argue that lexical priority relations have a different source – that a large enough number of headaches *would* have a greater total disvalue than a death, but we would be obliged to choose them nonetheless, because acting morally is not always a matter of minimising total disvalue. While decision theory is, I think, flexible enough to emulate the predictions of such views, the question of whether it can be interpreted in a way that is faithful to their motivations is not one that I take up here. For discussion of these issues see, for instance, Colyvan, Cox and Steele (2010), Brown (2011), Lazar (2017), Smith and Black (2019), Black (2020) and Lazar and Graham (2021).

According to *expected utility theory* (EUT), which will serve as our starting point, the risk function assigns *probabilities* to the outcomes that could result from an action – that is, positive real numbers that sum to 1 – while the value function assigns *utilities* to these outcomes – that is, positive or negative numbers. On a moral interpretation, the probabilities should be understood as epistemic or evidential, and the utilities should be understood as measuring objective value. The decision rule selects the action or actions that have the highest *expected utility*, where the expected utility of an action is equal to the average of the utilities of each of its possible outcomes, weighted by their probabilities. When two or more actions are tied for the highest expected utility, all of these actions are selected. Utilities, when negative, might be referred to as *disutilities* and, if it is potential negative outcomes that bulk large in a decision problem, we might say that the decision rule selects the action or actions that minimise expected disutility.

If \mathbf{A} is the set of available actions, \mathbf{O} is the set of possible outcomes, Pr is the probability function, and u is the utility function then, as is familiar, the expected utility of $A \in \mathbf{A}$ can be written like this:

$$EU(A) = \sum_{O \in \mathbf{O}} \text{Pr}(O | A) \cdot u(O)^4$$

On a moral interpretation, the action or actions that maximise expected utility will be morally permissible for an agent, while the rest will be morally prohibited. In a case where only a single action maximises expected utility, this action will be morally required.⁵

Following Jeffrey (1965), the A s and O s can be regarded as *propositions* – the propositions that the action in question is performed or that the outcome in question is realised. Let propositions be modelled as subsets of a set of possible worlds W and let Ω be a Boolean σ -algebra of propositions – a set of propositions that contains W and is closed under negation and disjunction (and under countable disjunction in case Ω is infinite). The set of actions \mathbf{A} and the set of outcomes \mathbf{O} will both be partitions of W that are subsets of Ω . We assume here that outcomes are individuated in such a way as to include the action that gives rise to them (see Colyvan, Cox and Steele 2010: section 3), in

⁴ The formula used here is characteristic of *evidential* decision theory. In *causal* decision theory the probability assigned to an outcome O , given an action A , is understood not as a conditional probability, but as a kind of ‘causal’ probability which, roughly speaking, reflects only A ’s *causal influence* upon O , bracketing any purely *evidential* connection between A and O (see Gibbard and Harper 1978; Lewis 1981; Joyce 1999; Egan 2007; Buchak 2016: section 2). The dispute between evidential and causal decision theory (on which I mean to take no position here) is orthogonal to the issues under discussion – in the sense that none of the examples I consider would prompt a different treatment from the two approaches. There are substantial questions about how one might develop ‘causal’ versions of the various decision theoretic frameworks that I canvass – but these will have to be postponed for another occasion.

⁵ For more on moral interpretations of expected utility theory see Broome (1991), Jackson (1991, 2001), Colyvan, Cox and Steele (2010), Lee-Stronach (2018: section 2), Lazar and Lee-Stronach (2019: section 2).

which case \mathbf{O} will be a refinement or fine-graining of \mathbf{A} (every element of \mathbf{O} is a subset of an element of \mathbf{A}).⁶

Pr can be defined as a function taking the members of Ω to real numbers in the unit interval and conforming to the probability axioms:

For $X, Y \in \Omega$:

$$P1 \quad \Pr(W) = 1$$

$$P2 \quad \Pr(X) \geq 0$$

$$P3 \quad \text{If } X \text{ and } Y \text{ are inconsistent then } \Pr(X \vee Y) = \Pr(X) + \Pr(Y)$$

If we are dealing with an infinite stock of propositions we might strengthen P3 to:

$$\text{If } \Pi \subseteq \Omega \text{ is a set of pairwise inconsistent propositions then } \Pr(\bigvee \Pi) = \sum_{X \in \Pi} \Pr(X)$$

The conditional probabilities required for calculating expected utilities can be derived according to the standard ratio formula: $\Pr(X | Y) = \Pr(X \wedge Y)/\Pr(Y)$, and is undefined in case $\Pr(Y) = 0$. Given this definition, we can easily prove a *conditional* version of the axiom P3:

$$CP3 \quad \text{If } X \text{ and } Y \text{ are inconsistent and } \Pr(Z) > 0 \text{ then } \Pr(X \vee Y | Z) = \Pr(X | Z) + \Pr(Y | Z).^7$$

If there is only a single outcome O associated with an action A , then A and O will be identical propositions and the expected utility of A will be equal to whatever utility is assigned to O . If all available actions are like this, we have a decision under certainty, in which $\mathbf{A} = \mathbf{O}$. Otherwise, we have a decision under risk.⁸ In a decision under certainty an agent is, in effect, choosing between outcomes,

⁶ As mentioned in n2 a formal model of a decision problem will sometimes include a set of *states*, which allow us to factor an outcome into a contribution due to the agent and a contribution due to the world. Formally, the set of states \mathbf{S} will be another partition of Ω , such that every outcome $O \in \mathbf{O}$ is associated with an action, state pair. The probability of an outcome O , given an action A will be equal to the sum of the probabilities, given A , of those states that would, in combination with A , lead to O . This additional structure allows us to define richer relations between actions – such as relations of dominance.

⁷ Assume X and Y are inconsistent and $\Pr(Z) > 0$.

$$\begin{aligned} \Pr(X \vee Y | Z) &= \Pr((X \vee Y) \wedge Z)/\Pr(Z) && [\text{Def}^n \text{ of conditional probability}] \\ &= \Pr((X \wedge Z) \vee (Y \wedge Z))/\Pr(Z) \\ &= (\Pr(X \wedge Z) + \Pr(Y \wedge Z))/\Pr(Z) && [P3] \\ &= \Pr(X \wedge Z)/\Pr(Z) + \Pr(Y \wedge Z)/\Pr(Z) \\ &= \Pr(X | Z) + \Pr(Y | Z) && [\text{Def}^n \text{ of conditional probability}] \end{aligned}$$

⁸ It is common in decision theory to distinguish between a decision under *risk*, in which it is possible to assign a probability to each outcome given each action, and a decision under *uncertainty* or *ignorance*, in which some probabilities cannot be assigned and we may wish to use a decision rule that selects actions based only on the utilities of their possible outcomes (see for instance Resnik 1987: p14, chap. 2; Joyce 1999: p16). This way of drawing the distinction presupposes, however, that risk is to be understood exclusively in probabilistic terms and, since I will be questioning this assumption, we might opt for something more neutral: A decision under risk is one in which the risk function is defined for each action/outcome pair, and a decision under uncertainty is one in which there are action/outcome pairs for which the risk function goes undefined. My attention here will be restricted to decisions under risk, so characterised.

based upon their utilities. Suppose O^* is an outcome in which an innocent person dies and O' , O'' , O''' ... are a series of outcomes in which one person suffers a mild headache, two people suffer a mild headache, three people suffer a mild headache and so on. If preserving the life of an innocent takes lexical priority over the avoidance of headaches then, in a decision under certainty, one is morally required to choose any one of the outcomes in O' , O'' , O''' ... over O^* . In EUT, this will require that O^* be assigned a disutility that is greater than every outcome in O' , O'' , O''' ... But if every headache adds a constant disutility⁹ then, given the constraints of the framework, the only way to achieve this is to allow O^* to be assigned a disutility that is *infinite* (Jackson and Smith 2006; Huemer 2010; Colyvan, Cox and Steele 2010: section 3).¹⁰

If O^* is assigned infinite disutility, then we can capture the claim that preserving the life of an innocent takes lexical priority over the avoidance of headaches. But this strategy leads to a number of untoward consequences. The trouble with an infinite disutility is that it *remains infinite* when multiplied by any positive probability (Lee-Stronach 2018: section 3, for related discussion see Hájek 2003: section 3). If O^* has an infinite disutility, and an action has a non-zero probability of resulting in O^* then one of the terms in the expected utility calculation for the action will be *negative infinity*. In

⁹ If we are prepared to accept that there is no minimal unit of utility or disutility, then headaches could be assigned a *diminishing marginal disutility* tending towards 0. In this case, the disutility of O' , O'' , O''' ... would tend to a finite limit and the death of an innocent could be assigned a disutility that exceeds this limit (Lazar and Lee-Stronach 2019). Suppose, for instance, that O' is assigned a utility of -1, O'' is assigned a utility of $-1\frac{1}{2}$, O''' is assigned a utility of $-1\frac{3}{4}$ and so on. If O^* were assigned a utility of -2, we would have the desired result that an agent would be morally required to choose any member of O' , O'' , O''' ..., rather than O^* . There are significant drawbacks, however, to accommodating lexical priority in this way. Most obviously, it requires us to place different value on the wellbeing of different people. I won't discuss this further here – but for more on the general strategy of capturing lexical priority by placing finite bounds on the utility and disutility of certain benefits and costs see Black (2020) and Hawthorne, Isaacs and Littlejohn (2023).

¹⁰ Suppose we consider all of the possible *lotteries* that can be constructed from a set of outcomes O – that is, the first order lotteries that confer probabilities upon the outcomes, the second order lotteries that confer probabilities upon the first order lotteries and the outcomes, the third order lotteries that confer probabilities upon the first and second order lotteries and the outcomes and so on. Provided an agent's preferences amongst these outcomes and lotteries conform to certain axioms, their preferences over O can be represented by a finite-valued utility function, unique up to positive linear transformation (see, for instance, Resnik 1987: chap. 4, Steele and Stefánsson 2016: section 2.3). If we accept that rational preferences must conform to these axioms, then we have an argument for limiting ourselves to finite utilities and disutilities. On a *moral* interpretation, in order for this kind of argument to engage, we would need to think of the axioms as representing constraints upon the comparative value predictions of any viable axiology.

One axiom that is crucial for driving this representation theorem is the Archimedean or *continuity* axiom. On the moral interpretation, the continuity axiom entails that if O_1 is better than O_2 which is in turn better than O_3 then, for some real number p , O_2 is just as good as a lottery which has a p probability of leading to O_1 and a $1-p$ probability of leading to O_3 . In effect, the continuity axiom guarantees that the value difference between any two outcomes can be bridged by lotteries that confer different probabilities upon them. On its face, it's not obvious that any viable value assignment would have to satisfy this constraint. Many would take the view that a lottery which risks a severe negative outcome should never be tolerated for a minor benefit, no matter the probabilities involved. Whatever the case, it is clear that any approach which allowed infinite utilities or disutilities and, thus, infinite value differences, would permit violations of the axiom (Colyvan, Cox and Steele 2010: pp513-515; Hájek 2018: n7). The significance of representation theorems is not my primary topic here, but I will return to the continuity axiom in n14 and n27.

this case, the other terms in the calculation (if finite) will make no difference to the result – the expected utility of the action will also be negative infinity.

Suppose a patient is suffering from a disease which will be fatal if left untreated, and a doctor must choose between three different medications to administer. While all three medications will cure the disease, each has a chance of triggering a fatal reaction in the patient; there's a 0.2 probability that the patient will have a fatal reaction to the first medication, a 0.01 probability that they will have a fatal reaction to the second and a 0.00000001 probability that they will have a fatal reaction to the third. All else is equal, it's clear that the doctor would be morally obliged to administer the third medication. More generally, it's clear that administering the third medication would be morally preferable to administering the second which would, in turn, be morally preferable to administering the first. And yet, on the infinite disutility strategy, the actions would all turn out to be *morally equivalent* – as each will have an expected utility of negative infinity. One might think this puts the doctor in a kind of a moral *dilemma*, in which all available actions are prohibited (see Hayenhjelm and Wolff 2012: section 7, Holm 2016). As Lee-Stronach points out, however, moral dilemmas are effectively precluded on a moral interpretation of expected utility theory, and the problem is really one of *permissiveness* (Lee-Stronach 2018: section 3). If every available action in a decision problem has the same expected utility – even if this happens to be negative infinity – then they will all be selected by the decision rule and deemed morally permissible.

The problem of permissiveness stems from the fact that an infinite disutility will *swamp* any expected utility calculation in which it is allowed to feature. This also leads to what Lee-Stronach dubs the problem of *risk* (Lee-Stronach 2018: section 4). If a death is assigned an infinite disutility, then an action that is *certain* to result in a death will have the same expected disutility – infinite – as an action that merely *risks* a death, even if the risk involved is very low. In some sense, the infinite disutility strategy deems certain death and a risk of death to be *equally (infinitely) bad*. As a result, one would be morally obliged to choose any number of headaches over an action that risks a death. But many would take the view that if the risk of death is very low, then it may be morally acceptable to choose the action – particularly if the risk is comparable to that involved in everyday activities such as driving (see for instance Norcross 1998). If a friend is feeling a headache coming on then many of us would be perfectly willing to take a short drive to the pharmacy to buy them some painkillers even though we are aware, at some level, that this action might result in an accident in which an innocent person loses their life. Even if the probability that the drive will result in a death is, say, 0.00000001 (1 in 10 million), on the infinite disutility strategy, the drive would still have an expected utility of negative infinity.¹¹

¹¹ Another troubling feature of an infinite disutility is that, unless we introduce multiple orders of infinity, it will represent an *upper limit* on the disutility of any possible outcome. If a death is assigned an infinite disutility, as the strategy demands, then there will be no scope to assign a greater disutility to any other outcomes – including, say, two deaths or three deaths etc. If one were forced into a choice between these outcomes then they would all be deemed morally permissible, as they would have the same (infinite) disutility. This might be regarded as a

The problem of permissiveness and the problem of risk seem closely related – and may appear to share the same source. But that’s not to say that they necessarily have the same *fix*. Following Lee-Stronach, I will be arguing that these two problems in fact require different amendments to the framework of expected utility theory, with the first prompting us to rethink the value function and the second prompting us to rethink the risk function.

3 Multidimensional Utility

To make headway with the problem of permissiveness, we need to move away from the idea that lexical priority requires us to posit *infinite* differences in value – to regard certain outcomes as being infinitely worse than others. An alternative approach is to suppose that certain factors can only contribute to the value or disvalue of an outcome on the condition that *other* factors are not in play. When comparing the value of two possible outcomes it may be that the number of headaches being suffered will only be a relevant consideration in so far as no *lives* are hanging in the balance. On this way of thinking, it’s not that a death is 1,000 times or 1,000,000 times or an infinite number of times as morally bad as a headache. It’s better to say that the two cannot be directly compared, because a headache will only assume moral significance when lives are not at stake – that is, when appropriate precautions have been taken against the loss of life.

This perspective might be captured in a framework known as ‘lexicographic’ expected utility theory (LEUT).¹² Rather than supposing that the values of outcomes can be measured along a single dimension, each outcome is now assigned a *multidimensional* utility – a sequence of positive or negative numbers. The possible outcomes in a decision model can then be placed in a lexicographic order, in much the same way that one would order words alphabetically; with respect to the first utility dimension (first letter), then with respect to the second utility dimension (second letter) in the case of any ties, then with respect to the third in the case of further ties and so on (thus ‘lexicographic’ – see Rawls 1971: n23). When comparing the value of two outcomes, the second dimension will only contribute to the comparison if the outcomes don’t differ with respect to the first dimension, and the third will only contribute to the comparison if the outcomes don’t differ with respect to the first or second and so on.¹³

third problem that besets the infinite utility strategy – albeit one that can be easily solved by the measures introduced in the next section. I will return to this in n15.

¹² The possibility of lexicographic decision theory was mooted by von Neumann and Morgenstern (1947: p631), and developed in detail by Hausner (1954) (see also Fishburn 1982; Blume, Brandenburger and Dekel 1989). This framework has not been widely discussed in the philosophical literature – but see Hájek, (2003; section 4.2), Lee-Stronach (2018: section 3) and Russell and Isaacs (2021: sections 4 and 5). Lee-Stronach, as mentioned, uses the framework for the very same purpose that I do here.

¹³ More formally, if the multidimensional utility of O_1 is $\langle u_1(O_1), u_2(O_1), u_3(O_1) \dots \rangle$ and the multidimensional utility of O_2 is $\langle u_1(O_2), u_2(O_2), u_3(O_2) \dots \rangle$ then the first outranks the second just in case, for some u_i , $u_i(O_1) > u_i(O_2)$ and for all $j < i$ $u_j(O_1) = u_j(O_2)$.

On this framework, the expected utility of an action will also be multidimensional, with the first dimension equal to the probability weighted average of the first utility dimensions of the outcomes that could result, the second dimension equal to the probability weighted average of the second dimensions etc. If the utility of an outcome O is written $\langle u_1(O), u_2(O), u_3(O) \dots \rangle$, then the expected utility of an action A can be written $\langle EU_1(A), EU_2(A), EU_3(A) \dots \rangle$ where

$$EU_i(A) = \sum_{O \in \mathcal{O}} Pr(O | A) \cdot u_i(O)$$

The available actions in a model of a decision problem can also be placed in a lexicographic order according to their multidimensional expected utilities, with the decision rule selecting whichever action or actions are ranked first.¹⁴ On a moral interpretation, these will be the actions that are morally permissible for an agent.

Rather than assigning a death a disutility that is infinitely greater than that of a headache, within LEUT we have a new option for realising lexical priority; the two can be assigned disutility relative to different *dimensions*. Suppose we posit just two dimensions to utility, such that a death is assigned the utility $\langle -1, 0 \rangle$ while a headache is assigned the utility $\langle 0, -1 \rangle$. In this case, the outcome O^* in which an innocent loses their life is assigned $\langle -1, 0 \rangle$ while the outcomes $O', O'', O''' \dots$ in which one person suffers a mild headache, two people suffer a mild headache, three people suffer a mild headache... will be assigned, respectively $\langle 0, -1 \rangle, \langle 0, -2 \rangle, \langle 0, -3 \rangle \dots$. In this case, O^* will rank lower than every outcome in O', O'', O''' and we will have the result that one is morally required, in a decision under certainty, to choose any one of these outcomes over O^* . Importantly, this is *not* because a death is reckoned to be infinitely worse than a headache – rather it is because a headache will only assume moral significance on the condition that appropriate precautions have been taken against the loss of life – and deliberately taking a life is inconsistent with this condition being met.

The lexicographic strategy avoids the problem of permissiveness. Suppose again that a doctor is forced to choose between administering three medications which have, respectively, a 0.2 probability, a 0.01 probability and a 0.0000001 probability of triggering a fatal reaction in the patient. Assuming that the options are otherwise equal, administering the first medication will have an expected utility of $\langle -0.2, 0 \rangle$, administering the second will have an expected utility of $\langle -0.01, 0 \rangle$ and administering the third will have an expected utility of $\langle -0.0000001, 0 \rangle$. If administering no

¹⁴ The *continuity axiom*, as mentioned in n6, is crucial to the transition from EUT to LEUT. Consider again the set of lotteries which can be constructed from the outcomes in \mathcal{O} . If we give up continuity, but maintain that an agent's preferences over these outcomes and lotteries conform to von Neumann and Morgenstern's other axioms (order and independence), their preferences over the outcomes can be represented by a lexicographic ordering of multidimensional utilities, where each dimension u_i is unique up to the summing of linear transformations of $u_1 \dots u_i$, in which u_i receives a positive linear transformation (see for instance, Fishburn, 1982: section 5.6; Blume, Brandenburger and Dekel 1989). If we are inclined to doubt that continuity represents a constraint on rational preferences or, on the moral interpretation, a constraint on legitimate value comparisons, then we have reason to adopt multidimensional over unidimensional utility.

medication has an expected utility of $\langle -1, 0 \rangle$ then we have the desired result that the doctor is morally required to choose the third.¹⁵

4 'Appropriate Precautions' and De Minimis Risk

While it offers a solution to the problem of permissiveness, LEUT is still subject to the problem of risk (see Lee-Stronach 2018; Hawthorne, Isaacs and Littlejohn 2023: section 1). If a death has the utility $\langle -1, 0 \rangle$ then an action such as driving to the pharmacy, which has a 0.00000001 probability of resulting in a death would, all else equal, have an expected utility of $\langle -0.00000001, 0 \rangle$. If letting one's friend suffer a headache has utility $\langle 0, -1 \rangle$ then this would outrank the drive to the pharmacy and would still be the preferred option in a choice between the two. On the lexicographic strategy, like the infinite utility strategy, one is still morally obliged to choose any number of headaches over an action that risks a death. Unlike the infinite utility strategy, this is not because a death is deemed an infinitely worse outcome that must be avoided at all costs. Rather, it is because a 0.00000001 probability of a death – indeed any positive probability of a death – will be enough, within LEUT, to neutralise the moral significance of a headache.

This observation already points the way to a possible solution. The lexicographic strategy, as discussed in the last section, is based on the idea that a headache will only assume moral significance on the condition that *appropriate precautions* have been taken against the loss of life. But what if we were to adopt a less demanding conception of what constitutes 'appropriate precautions'? In particular, what if taking appropriate precautions against the loss of life, and activating the moral significance of something like a headache, were compatible with there being some non-zero risk of a death? Suppose we specify a probability value to close to, but greater than, 0 to serve as the 'de minimis' threshold. If the probability that an outcome O would result from an action A is lower than t then O might be described as a *de minimis risk*. Suppose that, even if an action carries some risk of a death, if the risk is classified as de minimis, then the agent will still count as having taken appropriate precautions against the loss of life.

One way to implement this idea within LEUT would be to change the way that actions are ranked based on their expected multidimensional utilities – to allow an action with an expected utility of $\langle -0.00000001, 0 \rangle$ to somehow rank higher than actions with expected utilities of $\langle 0, -1 \rangle$, $\langle 0, -2 \rangle$, $\langle 0, -3 \rangle$... But the approach I pursue here is different, and intervenes at an earlier stage of the decision procedure. Rather than using the evidential probability function Pr to assign the weights in an expected utility calculation, we instead use a *truncated* function, which results from conditionalising Pr on the conjunction of the negations of every de minimis risk – every outcome which, given its associated action, has a probability less than or equal to the de minimis threshold t. If α is a function

¹⁵ The lexicographic strategy is also capable of dealing with the 'upper limit' problem discussed in n11. That is, we now have the resources to separate the disutility of a single death ($\langle -1, 0 \rangle$) from that of two deaths ($\langle -2, 0 \rangle$), three deaths ($\langle -3, 0 \rangle$) etc.

that maps each outcome to its associated action and φ is a proposition defined as $\wedge\{\sim O \mid O \in \mathbf{O} \wedge \Pr(O \mid \alpha(O)) < t\}$ then, for each utility dimension u_i , we have it that:

$$DEU_i(A) = \sum_{O \in \mathbf{O}} Pr_{\varphi}(O \mid A) \cdot u_i(O) = \sum_{O \in \mathbf{O}} Pr(O \mid A \wedge \varphi) \cdot u_i(O) \text{ }^{16}$$

Actions are then subject to the usual lexicographic ordering, but based upon their *de minimis multidimensional expected utilities*, as supplied by this formula, with the decision rule selecting the action or actions that rank highest. We might call this framework *de minimis lexicographic expected utility theory* (DLEUT).

Less formally, we might think of the calculation of de minimis multidimensional expected utilities as a two-step process: First, we list all of the outcomes that could possibly result from each action and discount any outcomes with a probability below t . Second, we update the probabilities of those outcomes that remain, and use these probabilities to calculate expected utilities, for each utility dimension, and generate a lexicographic ordering of actions in the usual way. On a moral interpretation of DLEUT, the action or actions that have the highest de minimis multidimensional expected utilities, calculated in this way, will be morally permissible for an agent. For present purposes, I don't commit to the stronger assumption that these are the *only* actions that are morally permissible. In particular, I leave it open whether the actions which maximise standard multidimensional expected utility (if different) might also be morally permitted. DLEUT embeds the idea that de minimis risks *may* be legitimately discounted for the purposes of decision making – but, on the present interpretation, it doesn't *require* that they be discounted.

If the outcome O^* in which an innocent loses their life is assigned utility $\langle -1, 0 \rangle$ while the outcomes $O', O'' \dots$ in which one person suffers a mild headache, two people suffer a mild headache ... are assigned, respectively, the utilities $\langle 0, -1 \rangle, \langle 0, -2 \rangle \dots$ then DLEUT will predict that one is morally obliged to choose any one of the latter over the former. When it comes to decisions under certainty DLEUT will make the same predictions as LEUT. But when it comes to decisions under *risk*, the predictions can diverge. Suppose again that driving to the pharmacy has a 0.0000001 probability of leading to a death. Suppose we set the de minimis threshold at 0.0000002 (1 in 5 million). In this case, even if a death is assigned a utility of $\langle -1, 0 \rangle$, this won't register in the de minimis expected utility of this action. Rather, driving to the pharmacy will have a de minimis expected utility of $\langle 0, 0 \rangle$ which will outrank the utility of letting one's friend suffer a headache ($\langle 0, -1 \rangle$) making it a morally permissible choice.¹⁷ DLEUT is capable of delivering the result that any number of headaches must be morally

¹⁶ Another option is to define proposition φ in terms of *states* (if we include them in our decision model) – $\wedge\{\sim S \mid S \in \mathbf{S} \wedge \Pr(S) < t\}$. This will lead to the same results provided we make the following assumptions: (i) Outcomes are associated with a unique state as well as a unique action – that is, \mathbf{O} is a more fine-grained partition than both \mathbf{A} and \mathbf{S} . (ii) States and actions are probabilistically independent – for any $S \in \mathbf{S}$ and $A \in \mathbf{A}$, $\Pr(S \mid A) = \Pr(S)$.

¹⁷ Note that DLEUT will not predict, on the present interpretation, that driving to the pharmacy is morally *obliged* since, when we switch to standard expected utility, the ranking of the two options is reversed. For this stronger result we would need an interpretation on which all *and only* the actions that maximise de minimis expected

preferred to a death, but *without* delivering the result that any number of headaches must be morally preferred to a mere risk of a death – not, at any rate, if the risk is sufficiently low.¹⁸

The strategy of introducing a de minimis threshold as a way of solving the problem of risk is also pursued by Lee-Stronach (2018: section 5). But, rather than moving to a new framework such as DLEUT, Lee-Stronach suggests that we should adjust the way in which we model decision problems – leaving out any possible outcomes that have a probability lower than the threshold. On Lee-Stronach's approach, the probability function that features in the model can no longer be interpreted as representing an agent's evidential probabilities, since there may be outcomes with a positive evidential probability that are nevertheless missing from the model. Rather, it would need to be interpreted along the lines of Pr_{φ} above – that is, as representing the agent's evidential probabilities conditionalised on the negations of all de minimis risks.

While Lee-Stronach's approach may in one sense lead to the same results, I think there are significant advantages to representing the discounting of outcomes in the decision rule itself, rather than treating it as something that precedes the very construction of a decision model. For one thing, the richer models deployed in DLEUT contain sufficient resources to construct a new, *updated* model in the event that the de minimis threshold is altered or the agent acquires new evidence. I consider these possibilities in turn.

Lee-Stronach accepts that the de minimis threshold is something that can vary from one context to another and may be sensitive to factors such the moral stakes and the time and resource pressures bearing on the decision-maker (Lee-Stronach 2018: pp800-801). I agree that the de minimis threshold can vary, and that these kinds of factors may have an influence on where it should be set. Changes in these factors may of course be reflected in the utility function or other aspects of a decision model, but they can also lead to changes in the de minimis threshold, which will have a distinctive effect on the moral permissibility of the actions one is considering. DLEUT allows us to model the effects of such changes such by inserting a new number into the definition of φ and recalculating the de minimis expected utilities of each candidate action.

Whether or not the de minimis threshold can shift in this way, it's clear that the acquisition of new *evidence* is something that can alter the moral permissibility of actions. DLEUT can be given a dynamic aspect based on the idea that an agent's evidential probability function will update by conditionalisation on any new evidence that the agent acquires. That is, if an agent with an evidential

utility are morally permissible for an agent. While there may be a case to be made for such an interpretation, it is not needed to solve the problem of risk.

¹⁸ It is often assumed that the discounting of low risk possibilities is something that is only useful for *bounded* agents, who have limited time and energy for making decisions (see for instance Adler 2007). But if the present considerations are right then taking every possibility seriously, no matter how low the risk, is not just something that can make a decision more laborious – it can actively *obscure* certain sources of value that would otherwise be revealed. If one is considering driving to the pharmacy, and treats it as a live possibility that the drive could result in a death then one will, in effect, treat this as a life-or-death decision which will, in turn, occlude the value of something like relieving a friend's headache. As a result, ignoring low risk outcomes may be something that is important even for an idealised agent (for related discussion see Smith 2022b: section 1).

probability function Pr acquires evidence E , their evidential probability function will shift to Pr_E , where $Pr_E(P) = Pr(P | E)$ for any $P \in \Omega$, and a new truncated evidential probability function $Pr_{E,\varphi}$ can then be calculated as above. This will provide a way of modelling changes in the moral permissibility of actions as evidence is acquired.

It is important to note that conditionalising the agent's original truncated evidential probability function Pr_φ on E will not, in general, lead to the same results as conditionalising Pr on E and constructing a new truncated function – that is, $Pr_{E,\varphi}$ will not in general be equal to $Pr_{\varphi,E}$. The reason, roughly put, is that conditionalising Pr on E can lead to a *reclassification* of de minimis risks – outcomes which counted as de minimis risks prior to the update may cease to do so once the update has taken place. This effect cannot be reproduced by directly conditionalising Pr_φ on E .¹⁹ In fact, a unique $Pr_{E,\varphi}$ cannot be derived, by any method, from Pr_φ and E alone – the nature of $Pr_{E,\varphi}$ will be sensitive to the Pr and t from which Pr_φ is generated.²⁰ As a result, any decision model that leaves these elements out, such as the models proposed by Lee-Stronach, will be too impoverished to capture these dynamics.

5 The Threat of Collapse

While it offers a solution to the problem of risk, introducing the notion of a de minimis risk brings a certain *instability* into the framework of expected utility theory. Suppose an agent is faced with a pile of 10 million indistinguishable pills. One of these pills is poisoned and will cause death if ingested. The remaining pills are painkillers that will relieve a headache that the agent's friend is currently suffering. The decision the agent faces is whether to hand over one of these pills for their friend to take. Suppose again that we set the de minimis threshold at 0.00000002. In this case the outcome in which the agent happens to hand over the one poisoned pill, with a probability of 0.00000001, will qualify as a de minimis risk. Even if a death is assigned a disutility of $\langle -1, 0 \rangle$ and a headache is assigned a disutility of $\langle 0, -1 \rangle$ DLEUT would appear to predict that it is morally permissible for the agent to proceed. Handing over a pill will have a de minimis expected utility of $\langle 0, 0 \rangle$ which exceeds the de minimis expected utility of letting a person suffer a headache $\langle 0, -1 \rangle$.

¹⁹ The potential reclassification of de minimis risks means that $Pr_{E,\varphi}$ may be impossible to reach, from Pr_φ , by conditionalisation on *any* proposition. After all, a reclassified de minimis risk will be assigned 0 probability by Pr_φ and a positive probability by $Pr_{E,\varphi}$ – a change which, as is well known, can never be induced by conditionalisation alone. Suppose, for instance, that an action A has three possible outcomes O_1 , O_2 and O_3 such that $Pr(O_1 | A) = 0.99$, $Pr(O_2 | A) = 0.0099$ and $Pr(O_3 | A) = 0.0001$. If we let $t = 0.001$, then O_3 will count as a de minimis risk in which case $Pr_\varphi(O_3 | A) = 0$. If we then learn $E = \sim O_1$ and conditionalise Pr on this proposition, we derive $Pr_E(O_2 | A) = 0.99$ and $Pr_E(O_3 | A) = 0.01$. With $t = 0.001$, O_3 no longer counts as a de minimis risk in which case $Pr_{E,\varphi}(O_3 | A) = 0.01$. Thus, there is no proposition X such that $Pr_{\varphi,X} = Pr_{E,\varphi}$.

²⁰ For a simple illustration let $Pr' = Pr_\varphi$ from the previous footnote. While $Pr'_\varphi = Pr_\varphi$, we have it that $Pr'_{E,\varphi} \neq Pr_{E,\varphi}$ (since $Pr'_{E,\varphi}(O_3 | A) = 0$ and $Pr_{E,\varphi}(O_3 | A) \neq 0$).

Now suppose that there are 10 poison pills in the pile. With this change, one might assume that the outcome in which the agent's friend is poisoned would no longer count as a *de minimis* risk, leading to a reversal of the above verdict. But matters are not so clear. Since there are now 10 separate poison pills, we effectively have a *choice* as to how we specify the possible outcomes of this action. Does handing over two distinct poison pills count as two distinct outcomes or as the same outcome? If we give the first answer then there will be 10 different poison pill outcomes, each of which, with a probability of 0.00000001, will be classified as a *de minimis* risk. In this case the *de minimis* expected utility of handing over a pill will remain at $\langle 0, 0 \rangle$ and DLEUT will continue to predict that handing over a pill is morally permissible. If we give the second answer, then there will only be one poison pill outcome which, with a probability of 0.0000001, will not be classified as a *de minimis* risk. In this case, the *de minimis* expected utility of handing over a pill will be $\langle -0.0000001, 0 \rangle$ which will be exceeded by the *de minimis* expected utility of letting the friend suffer a headache. Since the *de minimis* expected utility and standard expected utility will be equal in this case, DLEUT will predict that the agent is morally prohibited from handing over a pill. So which is it?

Neither LEUT nor EUT faces this problem. Irrespective of whether we associate a unique outcome with each poison pill, or roll these together into a single outcome, the standard expected utility of handing over a pill is unchanged. More generally, the predictions of LEUT and of EUT have the property of *invariance under fine-graining*. Suppose we have a coarse-grained model which includes an outcome $O_{1 \vee \dots \vee n}$ and a fine-grained model in which $O_{1 \vee \dots \vee n}$ is divided into O_1, \dots, O_n but which is otherwise identical. Suppose we stipulate that the utility of $O_{1 \vee \dots \vee n}$ in the coarse grained model is equal to the probability weighted average of the utilities of O_1, \dots, O_n in the fine-grained model – that is, $u_i(O_{1 \vee \dots \vee n}) = \sum_{1 \leq x \leq n} \Pr(O_x | O_{1 \vee \dots \vee n}) \cdot u_i'(O_x)$, where the u_i s are the utility dimensions in the coarse grained model and the u_i' s are the utility dimensions in the fine grained model. In this case the utility of $O_{1 \vee \dots \vee n}$ in the coarse-grained model is, in effect, equal to the expected utility of $O_1 \vee \dots \vee O_n$ in the fine-grained model. Given this constraint, the expected utilities of all actions will be the same in both models (see for instance Joyce 1999: p121; Smith 2022b: appendix A). But even if this condition is observed, substituting $O_{1 \vee \dots \vee n}$ and O_1, \dots, O_n can still make a difference to the *de minimis* expected utility of an action, as the above example illustrates.

At the very least, a defender of DLEUT is in need of some principled way of settling on a decision model with a particular fineness of grain. One initially tempting thought is that the 10 poisoned pill outcomes should be grouped together on the grounds that handing over any one of these pills would result in the *same* disutility. We could even propose a general policy on which we distinguish outcomes O_1 and O_2 in a decision model iff $u_i(O_1) \neq u_i(O_2)$ for some u_i . With this policy, DLEUT would give a definite verdict in the preceding case – namely, that it is impermissible to hand over a pill. But now imagine a variant of the case in which the 10 different pills are associated with *different* utilities; handing over the first pill would result in a single death ($\langle -1, 0 \rangle$), handing over the second pill would result in two deaths ($\langle -2, 0 \rangle$), handing over the third pill would result in three deaths ($\langle -3, 0 \rangle$) and so on. Under the current policy we would now be forced to differentiate these outcomes in our model, in which case they would once again be classified as *de minimis* risks. But we would then

have the absurd result that it is impermissible to hand over a pill in the original case, but *permissible* to hand over a pill in this new case, even though there are even more lives at stake.

Another policy we could adopt, when constructing a decision model, is to distinguish outcomes as *finely* as we possibly can. After all, a fine-grained model contains more information – information that is missing, or glossed over, in a coarse-grained model. A model in which the outcomes are differentiated more finely comes closer to Savage’s notion of a ‘grand world’ decision model in which every potentially relevant detail is included (see Savage 1954: chap. 5). It’s natural to think that the predictions of a fine-grained model ought to take precedence over those of a coarse-grained model, in case those predictions conflict (Joyce 1999: section 2.6; Thoma 2019: section V). If we adopted this policy then DLEUT would, once again, appear to yield a definite verdict in the preceding case – only this time it is the verdict that handing over a pill is morally *permissible*.

But we have not yet followed through on the policy of differentiating outcomes as finely as we can. In the same way that there are 10 different poison pills that the agent could hand over, there are 9,999,990 different *painkillers* that the agent could hand over. If the poison pills each get their own unique outcome in the model then so too should the painkillers. But we are now faced with a model in which the action of handing over a pill has 10 million separate outcomes each of which, with a probability of 0.00000001, is below the de minimis threshold. And this will prevent us from calculating *any* de minimis expected utility for the action. Let A be the action of handing over a pill and let $O_1, \dots, O_{10,000,000}$ be the fine-grained outcomes corresponding to each pill in the pile. We have it that $\Pr(O_1 | A) = \Pr(O_2 | A) = \dots = \Pr(O_{10,000,000} | A) < t$, in which case $\varphi = \sim O_1 \wedge \sim O_2 \wedge \dots \wedge \sim O_{10,000,000} = \sim A$ and $\sum_{O \in \mathcal{O}} \Pr_\varphi(O | A) \cdot u(O) = \sum_{O \in \mathcal{O}} \Pr(O | A \wedge \sim A) \cdot u(O)$, which is *undefined*. Rather than guiding us towards a definitive verdict about whether it is morally permissible to hand over a pill, the policy of finely differentiating outcomes has, in effect, caused the DLEUT framework to collapse.

More generally, assume that $\langle W, \Omega, \Pr \rangle$ is an infinite, atomless probability space – that is, assume that W contains an infinite number of worlds, Ω contains an infinite number of propositions and for any $X \in \Omega$, there is a $Y \in \Omega$ such that $Y \subset X$ and $\Pr(X) > \Pr(Y) > 0$. Given these assumptions, for any action A and any set of outcomes \mathcal{O} constructed from Ω there is a fine-graining \mathcal{O}' of \mathcal{O} such that for every $O_x \in \mathcal{O}'$, $\Pr(O_x | A) < t$ (see Smith 2010b; 2016: section 9.3).²¹ If t is the de minimis threshold then, relative to \mathcal{O}' , the de minimis expected utility of A will be undefined.

Lee-Stronach is aware that the use of a probabilistic de minimis threshold invites collapse, and proposes to replace it with what he calls an ‘odds-based’ threshold (Lee-Stronach 2018: p801). On this approach, an outcome will count as a de minimis risk if and only if its probability is low *relative* to some other outcome that could result from the action. For any outcome O if we let O_{\max} designate the outcome that is most likely, given $\alpha(O)$, then proposition φ could be reformulated along the following lines: $\wedge \{ \sim O | \Pr(O | \alpha(O)) / \Pr(O_{\max} | \alpha(O)) < t \}$. By this method we can avoid the eventuality in which

²¹ The proof of this result makes use of countable additivity (strengthened P3 from section 2) and Zorn’s Lemma. It also assumes that, for each $O_x \in \mathcal{O}$, $\Pr(O_x | A)$ is rational (can be expressed as r/k for r and k integers).

every outcome that could result from an action is classified as a de minimis risk, and we can ensure that the de minimis expected utility of an action will always be defined, no matter how finely its outcomes are individuated.²² But there is still a problem. No matter what decision model we adopt, there will always be the potential for further fine-graining to alter the probabilistic balance of the outcomes which can, in turn, alter the classification of de minimis risks and the de minimis expected utilities of the available actions.

If we assume, once again, an infinite and atomless probability space, we can prove a corollary of the above result: For any action A , any set of outcomes \mathcal{O} and any outcome $O_y \in \mathcal{O}$ such that $\Pr(O_y | A) > 0$ there is a fine-graining \mathcal{O}' of \mathcal{O} such that $O_y \in \mathcal{O}'$ and for every $O_x \in \mathcal{O}'$ such that $x \neq y$, $\Pr(O_x | A)/\Pr(O_y | A) < t$. If t is the de minimis threshold then every possible outcome of A , other than O_y , will be discounted for the purpose of calculating a de minimis expected utility. Since the process can be repeated for \mathcal{O}' , fine-graining will always have the potential to alter the de minimis expected utility of A (provided some of the possible outcomes of A are assigned different utilities). While the use of an odds-based de minimis threshold will protect against the collapse of the framework, it will not guide us to a stable set of predictions. I will return to this in section 7.

To summarise, the predictions of DLEUT, unlike those of EUT or LEUT, depend on how finely the outcomes in a decision model are individuated. As a result, a defender of DLEUT is obliged to adopt some principled policy regarding the individuation of outcomes. But what appears to be the most natural policy, that of individuating outcomes as finely as possible, will trigger the collapse of the framework, leaving the de minimis expected utilities of all actions undefined. And even if we adopt an odds-based threshold to guard against collapse, the predictions of DLEUT remain unstable under fine-graining. One might think that the present considerations add further momentum to the decision theoretic critique of lexical priority – for here is yet another failed attempt to capture lexical priority within a decision theoretic framework. As I will argue in the next section, however, the problems that beset DLEUT have nothing to do with lexical priority per se – rather, they come from asking too much from a single notion of *risk*.

6 Probabilistic and Normic Risk

In section 2 I noted that any decision theoretic framework requires a function that tracks the *risks* of each possible outcome, given each available action. So far we have assumed that the risk function works by assigning *probabilities* – the higher the probability the higher the risk, and the lower the probability the lower the risk. It has also been assumed, accordingly, that a de minimis threshold should take the form of a probability value, such as 0.00000002. Probabilities, on one level, are simply

²² Lee-Stronach's approach is inspired by the probabilistic acceptance rule defended by Lin and Kelly (2012a; 2012b) which, amongst other things, guarantees that the set of accepted propositions is always consistent. If we think of a de minimis risk as the negation of an accepted proposition, then this is equivalent to saying that the de minimis risks cannot, between them, exhaust all logical possibilities.

numbers in the unit interval that sum to 1 when all of the possible outcomes of an action are included. But this seemingly mundane constraint already has significant consequences for the *logic* of risk.

To adopt this ‘probabilistic’ account is to treat risk as a *quantity* – as something that can be divided and apportioned. On the probabilistic account, when an outcome is divided into an exclusive and exhaustive set of more specific outcomes, its risk is *also* divided – the risks of the fine-grained outcomes must add up to the risk of the original coarse-grained outcome. This is a direct result of CP3 from section 2. Suppose several people have contracted a contagious illness and have been placed in quarantine. If an agent were to walk into the quarantine facility, without taking any precautions, there is a high risk that they would be infected. But what is the risk that they would be infected in the first second after stepping through the door? On the probabilistic account, the risk is very low. If this scenario were repeated over and over it would be very rare for the agent to be infected during the first second. And the same goes for the next second, and the second after that etc. And yet, if the agent spends an hour in the facility and ends up infected, then this must have happened during *some* particular second.

Suppose the probability of the agent being infected, during their hour in the facility, is equal to 0.8. If we divide that hour into 3600 one second intervals then that 0.8 probability must be apportioned between them. And these one-second intervals could of course be *further* divided into intervals of a half second or a quarter second etc. in which case the probability will be spread even more thinly. On the probabilistic account, the high risk possibility of being infected during an hour in the facility can be divided up into a set of possibilities, each of which has an arbitrarily *low* risk. The divisibility of probabilistic risk is also illustrated by the poison pill example. Given the set-up, if the agent hands over a pill then there is a risk that their friend will be poisoned. But if we divide this outcome into 10 sub-outcomes based upon the 10 different poison pills then, on the probabilistic account, the risk must also be divided.

The divisibility of probabilistic risk is precisely what makes it apt for supplying the weights in an expected utility calculation – but it is fundamentally at odds with the idea of a *de minimis* risk. In the quarantine example, as we’ve seen, the probabilistic risk that the agent will be infected in the first second is very low. If the agent is entitled to discount possibilities with a very low risk then this may be something they are entitled to discount. But if the agent discounts the possibility of being infected in the first second then, to be consistent, they should also discount the possibility of being infected in the next second, and the second after etc., as these possibilities all have a similarly low risk. But if the agent discounts *every one* of these possibilities then they will in effect have discounted the possibility of being infected *at all*.²³

If the probabilistic account were the only legitimate way to think about risk, then the idea of a *de minimis* risk may need to be abandoned – which would have repercussions for the viability of lexical priority. But a number of authors have recently challenged the idea that risk must be thought of in probabilistic terms, and put forward alternatives such as the modal account (Pritchard 2015), the

²³ For further objections to the discounting of low probability outcomes see Kosonen (2024).

relevant alternatives account (Gardiner 2021) and the *normic* account (Ebert, Smith and Durbach 2021; Smith 2022b; Mace and O’Sullivan 2024) – which will be my focus here. On the normic account, the risk of a given outcome is determined by how *abnormal* it would be – the less abnormal the outcome the higher the risk, and the more abnormal the outcome the lower the risk.

Suppose again that an agent is considering whether to drive to the pharmacy to buy painkillers for a friend. The agent has the evidence that the pharmacy is only a short drive away, that the car has recently been serviced and is working fine, that they’re not especially tired or under the influence of alcohol or drugs etc. On both the probabilistic and normic accounts, the risk that this drive will result in a fatal accident will be rated as very low. On the probabilistic account the reason for this is that occurrences of fatal accidents on short drives with working cars and sober drivers etc. are very *infrequent* – such drives happen all the time and rarely result in fatal accidents. On the normic account the reason is somewhat different; if a short drive with a working car and sober driver *did* result in a fatal accident, then there would have to be some *explanation* as to how this could have possibly happened. Perhaps the driver suddenly lost consciousness or the brakes had been tampered with, or someone ran out right in front of the car etc. Whatever the case, a fatal accident is not a normal outcome of a decision to drive to the pharmacy – it is an outcome that requires significant explanation.²⁴ If we were to imagine instead that the driver is drowsy, drunk etc. then, on both accounts, the risk would be higher. Under these new conditions, fatal accidents would be more frequent and the need for explanation, in the event of a fatal accident, would be reduced.

The probabilistic and normic accounts will often be in broad agreement about individual risk assessments – but they don’t invariably agree, and they make quite different predictions about the logical properties of risk. Return to the quarantine example. Given that the new disease is contagious and the agent has taken no precautions, no special explanation would be needed if they were infected during their one hour in the quarantine facility. But the *same is true* of the first second and of the next second etc. None of these events would be abnormal in the sense of requiring special explanation. If the agent was wearing full PPE then that might introduce the need for special explanation in the event that they end up infected – perhaps the equipment is faulty or damaged or the virus survived on their clothing after they left the facility etc. – but simply narrowing in on a smaller time interval is not going to have that effect.

On the normic account there is a high risk that the agent will be infected during their one hour in the facility, but there is *also* a high risk that the agent will be infected in the first second, or the next second etc. If the agent thinks to themselves ‘There’s a low risk that I’ll be infected in the next second’ then on the probabilistic account that may be true, but on the normic account it’s false. In the poison pill example if the agent decides to hand over a pill, there is a high normic risk that their friend will be poisoned. If the pile contains 10 poison pills and the agent chooses at random then no special explanation is needed in the event that they hand over a poisoned pill. But the very same remarks apply even if we focus on just *one* of the poison pills. If the pill is somewhere in the pile and the agent

²⁴ For more on the link between normalcy and the need for explanation see Smith (2010a; 2016, chap. 2; 2022a).

chooses at random then no special explanation is needed in the event that the agent chooses that very pill.

Normic risk, unlike probabilistic risk, is not divisible. When an outcome is divided into a series of more specific outcomes, the normic risk of these new fine-grained outcomes may be *just as high* as the normic risk of the original coarse-grained outcome. If the de minimis threshold were defined in normic rather than probabilistic terms then the agent would not be entitled to ignore the possibility of being infected during the first second in the facility, or the possibility of being infected during the next second and so on. More generally, if the de minimis threshold is defined in normic terms, it will not be possible to circumvent that threshold by dividing up a coarse-grained outcome – no matter how finely we do it.

7 Multidimensional Risk

Rather than supposing that the risk of an outcome can be measured along a single dimension, in the final decision theoretic framework I will consider, every outcome is assigned two numbers by the risk function; one representing its *probability* and the other representing its *abnormality*, given an agent's evidence. As above, let Ω be a Boolean σ -algebra of propositions built from the elements of a set of possible worlds W . Suppose the propositions in Ω can be placed in an abnormality ordering, reflecting how much explanation their truth would require, given the evidence. Suppose that any two propositions can be compared for their normalcy – that is, suppose that, for any two propositions, either one is more normal than the other or both are equally normal. Suppose, in addition, that there are no infinite ascending chains of increasingly normal propositions that continue without end. In this case, the propositions in Ω could be assigned numerical abnormality degrees – the maximally normal propositions will be assigned a degree of 0, the next most normal propositions will be assigned a degree of 1 and so on.

No proposition can be more normal than the universal proposition W which will be assigned an abnormality degree of 0. No proposition can be less normal than the empty proposition \emptyset which will be assigned an infinite degree of abnormality. The final constraint on an abnormality ordering concerns the relation between disjunctions and their disjuncts: The only way in which $X \vee Y$ can be true is if either X is true or Y is true. To explain the truth of $X \vee Y$ one must either explain the truth of X or the truth of Y . As a result, the amount of explanation demanded by the truth of $X \vee Y$ will be equal to either the amount demanded by X or by Y – whichever is less. And the degree of abnormality of $X \vee Y$ will be equal to the degree of abnormality of X or of Y – whichever is lower (see Smith, 2022a, section 5, 2022b, section 4).

We can now define an evidential abnormality function ab taking members of Ω into the set of nonnegative integers plus infinity – $\{1, 2, 3, 4 \dots \infty\}$. Given the above constraints, ab will satisfy the axioms for a *negative ranking function* (see, for instance, Huber 2009: section 4; Spohn 2012: chap. 5; Smith 2016: chap. 8, n7; 2022b, section 4):

- R1 $ab(W) = 0$
 R2 $ab(\emptyset) = \infty$
 R3 $ab(X \vee Y) = \min\{ab(X), ab(Y)\}$

If we are dealing with an infinite stock of propositions we might strengthen R3 to:

$$\text{If } \Pi \subseteq \Omega \text{ then } ab(\vee \Pi) = \min\{ab(X) \mid X \in \Pi\}$$

where Π is any set of propositions.

Conditional degrees of abnormality are defined by the following formula: $ab(X \mid Y) = ab(X \wedge Y) - ab(Y)$ (where $\infty - \infty = 0$) (Huber 2009: p19; Spohn 2012: section 5.3; Smith 2022b: section 4). In this case, the abnormality of X given Y is equal to the abnormality that X adds to the abnormality of Y – the amount of additional explanation that $X \wedge Y$ requires over and above that required by Y . This is equivalent to saying that the abnormality of Y and the abnormality of X given Y add up to the abnormality of $X \wedge Y$. Given this definition, we can easily prove a *conditional* version of the axiom R3:

$$\text{CR3} \quad ab(X \vee Y \mid Z) = \min\{ab(X \mid Z), ab(Y \mid Z)\}^{25}$$

On the normic account of risk, the risk that an outcome O will result from an action A can be gauged by the conditional abnormality of O given A – by $ab(O \mid A)$. The lower the value of $ab(O \mid A)$, the greater the risk that O will result from A , with the risk being maximal when $ab(O \mid A) = 0$ – indicating that O is one of the normal outcomes of A .

Suppose we specify some abnormality rank t to serve as the de minimis threshold. Rather than using Pr to assign the weights in an expected utility calculation, we now use a truncated function, which results from conditionalising Pr on the conjunction of the negations of every de minimis risk – every outcome which, given its associated action, has an abnormality greater than t . If α is once again a function that maps each outcome to its associated action and ψ is a proposition defined as $\wedge\{\sim O \mid O \in \mathbf{O} \wedge ab(O \mid \alpha(O)) > t\}$ then for each utility dimension u_i we have it that:

$$NDEU_i(A) = \sum_{O \in \mathbf{O}} Pr_{\psi}(O \mid A) \cdot u_i(O) = \sum_{O \in \mathbf{O}} Pr(O \mid A \wedge \psi) \cdot u_i(O)^{26}$$

Actions are then subject to a lexicographic ordering, based upon their *normic de minimis multidimensional expected utilities*, as supplied by this formula, with the decision rule selecting the

²⁵ $ab(X \vee Y \mid Z) = ab((X \vee Y) \wedge Z) - ab(Z)$ [Defⁿ of conditional abnormality]
 $= ab((X \wedge Z) \vee (Y \wedge Z)) - ab(Z)$
 $= \min\{ab(X \wedge Z), ab(Y \wedge Z)\} - ab(Z)$ [R3]
 $= \min\{ab(X \wedge Z) - ab(Z), ab(Y \wedge Z) - ab(Z)\}$
 $= \min\{ab(X \mid Z), ab(Y \mid Z)\}$ [Defⁿ of conditional abnormality]

²⁶ As with DLEUT, we could opt to apply the de minimis threshold directly to states, defining ψ as $\wedge\{\sim S \mid S \in \mathbf{S} \wedge ab(S) > t\}$. This will lead to the same results on the assumption that every outcome is associated with a unique state, and states and actions are *normically* independent – for any $S \in \mathbf{S}$ and $A \in \mathbf{A}$, $ab(S \mid A) = ab(S)$.

action or actions that rank highest. We might call this framework *normic de minimis lexicographic expected utility theory* (NDLEUT). On a moral interpretation of NDLEUT, the action or actions that have the highest normic de minimis multidimensional expected utility will be morally permissible for an agent. As with DLEUT, I won't commit to the stronger interpretation on which these are the only actions which are morally permissible.

If the outcome O^* in which an innocent loses their life is assigned utility $\langle -1, 0 \rangle$ while the outcomes $O', O'' \dots$ in which one person suffers a mild headache, two people suffer a mild headache, ... are assigned, respectively, the utilities $\langle 0, -1 \rangle, \langle 0, -2 \rangle \dots$ then NDLEUT will predict that one is morally obliged to choose any one of the latter over the former. In a decision under certainty the predictions of NDLEUT will converge with those of DLEUT and LEUT – it is only when it comes to decisions under risk that the distinctive features of the three frameworks emerge.

In the choice over whether to drive to the pharmacy to buy painkillers for a friend, NDLEUT offers the same predictions as DLEUT. As observed in the last section, it would not only be highly *improbable* for a drive to the pharmacy to result in a fatal car accident, it would also be highly *abnormal*. That is, this outcome could be regarded as a de minimis risk on either a probabilistic or a normic construal. Given a suitable de minimis threshold, even if a death is assigned a utility of $\langle -1, 0 \rangle$, the act of driving to the pharmacy will have a normic de minimis expected utility of $\langle 0, 0 \rangle$ which will outrank the utility of letting one's friend suffer a headache ($\langle 0, -1 \rangle$). NDLEUT, like DLEUT, offers a solution to the problem of risk; just because any number of headaches must be morally preferred to a death, it doesn't follow that any number of headaches must be morally preferred to a risk of death – not if the risk is sufficiently low.

NDLEUT avoids the threat of collapse. Return to the poison pill example. In this case, an agent is confronted with a pile of 10 million indistinguishable pills, 10 of which are poisoned, while the remainder are painkillers. Suppose again that a death is assigned a utility of $\langle -1, 0 \rangle$ and a headache is assigned a utility of $\langle 0, -1 \rangle$. As discussed in section 5, if we adopt a fine-grained model in which we have one outcome for every pill in the pile then, within DLEUT, every outcome of handing over a pill could count as a de minimis risk, and we will be unable to calculate a de minimis expected utility for the action. NDLEUT avoids this result. If the pill is chosen at random then for the agent to choose any one pill requires no more explanation than the agent choosing any other.

Let A be the action of handing over a pill and let $O_1, \dots, O_{10,000,000}$ be the fine-grained outcomes of this action – one for each pill in the pile. We have it that $ab(O_1 \vee \dots \vee O_{10,000,000} \mid A) = 0$. If the outcomes are all equally normal then $ab(O_1 \mid A) = \dots = ab(O_{10,000,000} \mid A)$. It now follows, by CR3, that the outcomes are all maximally normal – $ab(O_1 \mid A) = \dots = ab(O_{10,000,000} \mid A) = 0$ in which case none of them will count as a de minimis risk, no matter where the threshold is set. Handing over a pill has a normic de minimis expected utility of $\langle -0.000001, 0 \rangle$ which will be exceeded by the normic de minimis expected utility of letting a person suffer a headache ($\langle 0, -1 \rangle$) in which case, according to NDLEUT, it will be impermissible for the agent to proceed. This prediction will, in fact, hold true irrespective of how many poison pills and how many painkillers there are in the pile – whether it is 10 out of 10

million, 1 out of 10 million, 1 out of a billion etc. As long as there are some poison pills in the pile, and the pill is selected at random, the outcome in which the agent hands over a poison pill will never qualify, within NDLEUT, as a de minimis risk.²⁷

More generally, there are no NDLEUT decision models in which every possible outcome of an action is classified as a de minimis risk. Suppose O_1, \dots, O_n are all of the possible outcomes of an action A . In this case we have it that $A \wedge (O_1 \vee \dots \vee O_n) = A$ in which case $ab(A \wedge (O_1 \vee \dots \vee O_n)) = ab(A)$. By the definition of conditional abnormality $ab(O_1 \vee \dots \vee O_n | A) = ab(A \wedge (O_1 \vee \dots \vee O_n)) - ab(A) = 0$. It then follows, by CR3, that there is at least one $O_x \in \{O_1, \dots, O_n\}$ such that $ab(O_x | A) = 0$ in which case O_x cannot count as a de minimis risk and the normic de minimis expected utility of A will be defined. As a result, a policy of individuating outcomes as finely as possible will never cause NDLEUT to collapse – no matter how fine-grained the model, NDLEUT will continue to make concrete predictions. It's important to note, though, that NDLEUT can make *different* predictions with respect to a fine-grained and a coarse-grained model. That is, the predictions of NDLEUT, unlike those of EUT or LEUT, are not completely invariant under fine-graining.

Suppose we have a coarse-grained model which includes an outcome $O_{1 \vee \dots \vee n}$ and a fine-grained model in which $O_{1 \vee \dots \vee n}$ is divided into O_1, \dots, O_n but is otherwise identical. Recall that, provided the utility of $O_{1 \vee \dots \vee n}$ in the coarse grained model is equal to the expected utility of $O_1 \vee \dots \vee O_n$ in the fine-grained model – $u_i(O_{1 \vee \dots \vee n}) = \sum_{1 \leq x \leq n} \Pr(O_x | O_{1 \vee \dots \vee n}) \cdot u_i'(O_x)$ – the expected utility of A will be the same in both models. Suppose that $ab(O_1 \vee \dots \vee O_n | A) \leq t$ in which case $O_{1 \vee \dots \vee n}$ is not a de minimis risk in the coarse-grained model. CR3 guarantees that there must be at least one outcome $O_x \in$

²⁷ This prediction is in direct conflict with the *continuity axiom* mentioned in n10. If O_1 is better than O_2 which is in turn better than O_3 then the continuity axiom requires that, for some real number p , O_2 is just as good as a lottery which has a p probability of leading to O_1 and a $1-p$ probability of leading to O_3 . If we let O_3 be the outcome in which my friend is poisoned, O_2 be the outcome in which my friend suffers a headache and O_1 be the outcome in which my friend is headache free, we have our desired counterexample – O_2 will be preferable to any lottery that has O_3 as an outcome. As explained in n10, continuity is crucial in proving that a comparative value ordering of outcomes can be represented by a unidimensional utility function, and it is no surprise that this axiom should fail for any framework that incorporates a lexicographic element. While the 'normic de minimis' element of the framework is not instrumental to the failure of the axiom, it can perhaps offer a *rationale* as to why the axiom should be discarded.

As remarked in n10, many would take the view that a minor benefit is never worth a lottery that risks a severe negative outcome, no matter the probability of that outcome (Temkin 2001). Defenders of the continuity axiom, and of EUT, typically try to mollify our intuitions about such cases by alleging that we are regularly involved in lotteries of this kind, and willing to accept them without a second thought (Steele and Stefánsson 2016: section 2.3; Arrhenius and Rabinowicz 2005: p179, see also see also Joyce 1999: p94). Here is one example of such reasoning due to Steele and Stefánsson: 'Is there any probability p such that you would be willing to accept a gamble that has that probability of you losing your life and probability $(1-p)$ of you winning \$10? Many people think there is not. However, the very same people would presumably cross the street to pick up a \$10 bill they had dropped. But that is just taking a gamble that has a very small probability of being killed by a car but a much higher probability of gaining \$10'. From the viewpoint of normic risk, the comparison is specious – the two choices are not equivalent, and crossing the street in the latter case in no way commits one to taking the gamble in the former. It is true, of course, that the difference between these two decisions cannot be represented within EUT – but to appeal to EUT in defending the continuity axiom would, in effect, be to argue in a circle.

$\{O_1, \dots, O_n\}$ such that $ab(O_x | A) \leq t$, but is consistent with there being another outcome $O_y \in \{O_1, \dots, O_n\}$ such that $ab(O_y | A) > t$. That is if $O_1 \vee \dots \vee O_n$ is not classified as a de minimis risk in the coarse-grained model, CR3 is consistent with some, but not all, of the outcomes in $\{O_1, \dots, O_n\}$ being classified as de minimis risks in the fine-grained model. As a result, A may have a different normic de minimis expected utility in the two models, even if the expected utility constraint above is observed.

Importantly, though, it can be shown that there is a point in the process of fine-graining at which the predictions of NDLEUT will effectively *stabilise*. More precisely, it can be shown that, for any NDLEUT decision model, there is a more fine-grained model for which the normic de minimis expected utilities of the available actions will not be affected by any further fine-graining. The proof of this result is provided in the appendix. A policy of individuating outcomes as finely as possible will not lead to the collapse of NDLEUT – it will lead, rather, to a stable set of predictions.

In section 5, I discussed the possibility of using an *odds-based* threshold, rather than an absolute probability threshold, for the classification of de minimis risks. An odds-based variant of DLEUT will be collapse-proof, in the sense that the de minimis expected utilities of the available actions will always be defined, no matter how fine-grained the outcomes. However, for the reasons outlined in section 5, the use of an odds-based threshold will not secure the stability property just described. Given appropriate assumptions about the underlying probability space, for any odds-based DLEUT model there will be a more fine-grained model in which the available actions have different de minimis expected utilities.

Two final points. First, NDLEUT can also be used to model changes in moral permissibility prompted by the acquisition of new evidence. When an agent acquires a new piece of evidence E, we might suppose that the agent's evidential probability function Pr and the agent's evidential abnormality function ab will both be updated by conditionalisation on E. That is, Pr will shift to Pr_E , where $Pr_E(P) = Pr(P | E)$ for any $P \in \Omega$ and ab will shift to ab_E , where $ab_E(P) = ab(P | E)$ for any $P \in \Omega$. A new truncated evidential probability function $Pr_{E,\psi}$, used for the calculation of normic de minimis expected utilities, can then be constructed in the way described above. By this method, the acquisition of new evidence can lead to a change in the class of de minimis risks, with outcomes either gaining or losing de minimis status – $ab(O | A) \leq t$ is consistent with $ab_E(O | A) > t$ and $ab(O | A) > t$ is consistent with $ab_E(O | A) \leq t$. As was observed in the case of DLEUT, there is no shortcut to this new truncated function – $Pr_{E,\psi}$ will not in general be equivalent to conditionalising Pr_ψ on E. I hope to explore the dynamic aspect of NDLEUT in future work.

The second point concerns the logical independence of normic and probabilistic risk. As I have discussed, it is possible for an outcome to combine a low probabilistic risk with a high normic risk. This possibility is, in fact, crucial for ensuring that NDLEUT has the stability property described above. But the flipside of this is that it's also possible for an outcome to combine a high probabilistic risk with a low normic risk (one such case is described in Smith, 2022b, section 5). As a result, it is possible, within NDLEUT, for a high probability outcome to be classified as a de minimis risk. One could, in practice, avoid this result by adjusting the de minimis threshold to make it more demanding, particularly in a

case in which the stakes are high. As discussed in section 5 I regard a de minimis threshold as something that can vary from one context to another, in response to factors such as moral stakes. And even if a high probability outcome *were* classified as a de minimis risk, on the present interpretation of NDLEUT, that would mean only that one is entitled to ignore it, and not that one is *obliged* to ignore it. Having said that, the prospect of a high probability de minimis risk remains troubling, and a full treatment of this issue will have to await another occasion.

8 Conclusion

In this paper I have developed a new decision theoretic framework by gradually augmenting standard expected utility theory – first adding the ‘lexicographic’, then the ‘de minimis’ and finally the ‘normic’. I have argued that the resulting framework – NDLEUT – is capable of faithfully modelling relations of lexical priority while avoiding all of the main problems that have been identified in the literature. More than this, I have tried to show that these additions to expected utility theory are not just convenient technical devices, but can be squared with independently motivated conceptions of lexical priority and of risk.

And what, then, of the ‘decision theoretic’ critique of lexical priority – the claim that lexical priority becomes incoherent once we consider decisions under risk? Nothing that I have said here definitively shows that no critique of this kind could ever succeed. It could be that any framework capable of modelling lexical priority – including NDLEUT – is subject to insurmountable problems in the end (one potential problem for NDLEUT was just noted). But those who have pursued the decision theoretic critique have not, as yet, established any such conclusion. In its present form, the critique does more to expose the limitations of a particular decision theoretic framework than to cast doubt upon the idea of lexical priority.

Acknowledgements

Ideas connected with this material were presented at the University of Seville in February 2020, the University of St Andrews in October 2020, the University of Edinburgh in March 2022 and the University of Glasgow in September 2022. Thanks to everyone who participated on these occasions, and to two anonymous referees for this journal. Work on this paper was supported by the Arts and Humanities Research Council (grant no. AH/T002638/1).

References

- Adler, Matthew (2007) ‘Why de minimis?’ *Faculty Scholarship Paper* 158. Retrieved from http://scholarship.law.upenn.edu/faculty_scholarship/158
- Arrhenius, Gustav and Wlodek Rabinowicz (2005) ‘Value and unacceptable risk’ *Economics and Philosophy* v21, pp177-197

- Black, D (2020) 'Absolute prohibitions under risk' *Philosophers Imprint* v20(20), pp1-26
- Blume, Lawrence, Adam Brandenburger and Eddie Dekel (1989) 'An overview of lexicographic choice under uncertainty' *Annals of Operations Research* v19, pp231-246
- Brennan, Samantha (2006) 'Moral lumps' *Ethical Theory and Moral Practice* v9(3), pp249-263
- Broome, John (1991) 'The structure of good: Decision theory and ethics' in Bacharach, M. and Hurley, S. eds. *Foundations of Decision Theory* (Oxford: Blackwell)
- Buchak, Lara (2016) 'Decision theory' in Hájek, A. and Hitchcock, C. eds. *Oxford Handbook of Probability and Philosophy* (Oxford: Oxford University Press)
- Colyvan, Mark, Damian Cox and Katie Steele (2010) 'Modelling the moral dimension of decisions' *Noûs* v44(3), pp503-529
- Dorsey, Dale (2009) 'Headaches, lives and value' *Utilitas* v21(1), pp36-58
- Dougherty, Tom (2013) 'Aggregation, beneficence and chance' *Journal of Ethics and Social Philosophy* v7(2), pp1-19
- Dworkin, Ronald (1977) *Taking Rights Seriously* (London: Duckworth)
- Ebert, Philip, Martin Smith and Ian Durbach (2020) 'Varieties of risk' *Philosophy and Phenomenological Research* v101(2), pp432-455
- Egan, Andy (2007) 'Some counterexamples to causal decision theory' *Philosophical Review* v116(1), pp93-114
- Fishburn, Peter (1982) *The Foundations of Expected Utility* (Dordrecht: Reidel)
- Gardiner, Georgi (2021) 'Relevance and risk: How the relevant alternatives framework models the epistemology of risk' *Synthese* v199(1-2), pp481-511
- Gibbard, Alan and William Harper (1978) 'Counterfactuals and two kinds of expected utility' in *University of Western Ontario Series in Philosophy of Science* v15, pp153-190
- Hájek, Alan (2003) 'Waging war on Pascal's wager' *Philosophical Review* v112(1), pp27-56
- Hájek, Alan (2018) 'Pascal's wager' in Zalta, E. ed *Stanford Encyclopedia of Philosophy* (Summer 2018 Edition) URL = <https://plato.stanford.edu/archives/sum2018/entries/pascal-wager/>
- Hansson, Sven Ove (2013) *The Ethics of Risk: Ethical Analysis in an Uncertain World* (Basingstoke: Palgrave Macmillan)
- Hausner, Melvin (1954) 'Multidimensional utilities' in Thrall, R., Coombs, D. and Davis, R. *Decision Processes* (New York: Wiley)
- Hawthorne, John, Yoav Isaacs and Clayton Littlejohn (2023) 'Absolutism and its limits' *Journal of Moral Philosophy* v21(1-2), pp170-189

- Hayenhjelm, Madeleine and Jonathan Wolff (2012) 'The moral problem of risk impositions: A survey' *European Journal of Philosophy* v20(S1), e26-e51
- Holm, Sune (2016) 'A right against risk imposition and the problem of paralysis' *Ethical Theory and Moral Practice* v19(4), pp917-930
- Huber, Franz (2009) 'Belief and degrees of belief' in Huber, F. and Schmidt-Petri, C. eds. *Degrees of Belief* (Dordrecht: Springer)
- Huemer, Michael (2010) 'Lexical priority and the problem of risk' *Pacific Philosophical Quarterly* v91(3), pp332-351
- Jackson, Frank (1991) 'Decision-theoretic consequentialism and the nearest and dearest objection', *Ethics* v101(3), pp461-482
- Jackson, Frank (2001) 'How decision theory illuminates assignments of moral responsibility' in Naffine, N., Owens, R. and Williams, J. eds. *Intention in Law and Philosophy* (Aldershot: Ashgate)
- Jackson, Frank and Michael Smith (2006) 'Absolutist moral theories and uncertainty' *Journal of Philosophy* v103(6), pp267-283
- Jackson, Frank and Michael Smith (2016) 'The implementation problem for deontology' in Errol Lord and Barry Maguire eds. *Weighing Reasons* (Oxford: Oxford University Press)
- Jeffrey, Richard (1965) *The Logic of Decision* (New York: McGraw-Hill)
- Joyce, James (1999) *The Foundations of Causal Decision Theory* (Cambridge: Cambridge University Press)
- Kant, Immanuel (1797) 'On a supposed right to lie from philanthropic concerns' reprinted in Abbot, T. trans. *Kant's Critique of Practical Reason and Other Works on the Theory of Ethics* (London: Longman, Green and co., 1898)
- Kirkpatrick, James (2018) 'Permissibility and the aggregation of risks' *Utilitas* v30(1), pp107-119
- Korsgaard, Christine (1986) 'The right to lie: Kant on dealing with evil' *Philosophy and Public Affairs* v15(4), pp325-349
- Kosonen, Petra (2024) 'Probability discounting and money pumps' *Philosophy and Phenomenological Research* online first: <https://onlinelibrary.wiley.com/doi/abs/10.1111/phpr.13053>
- Lazar, Seth (2017) 'Deontological decision theory and agent centered options' *Ethics* v127(3), pp579-609
- Lazar, Seth and Chad Lee-Stronach (2019) 'Axiological absolutism and risk' *Noûs* v53(1), pp97-113
- Lazar, Seth and Peter Graham (2021) 'Deontological decision theory and lesser evil options' *Synthese* v198(7), pp6889-6916
- Lee-Stronach, Chad (2018) 'Moral priorities under risk' *Canadian Journal of Philosophy* v48(6), pp793-811

- Lewis, David (1981) 'Causal decision theory' *Australasian Journal of Philosophy* v59(1), pp5-30
- Lin, Hanti and Kevin Kelly (2012a) 'A geo-logical solution to the lottery paradox, with applications to conditional logic' *Synthese* v186(2), pp531-575
- Lin, Hanti and Kevin Kelly (2012b) 'Propositional reasoning that tracks probabilistic reasoning' *Journal of Philosophical Logic* v41(6), pp957-981
- Mace, Lilith and Angela O'Sullivan (2024) 'Reverse-engineering risk' *Erkenntnis* online first: <https://link.springer.com/article/10.1007/s10670-024-00788-6>
- Norcross, Alastair (1997) 'Comparing harms: Headaches and human lives' *Philosophy and Public Affairs* v26(2), pp135-167
- Norcross, Alastair (1998) 'Great harms from small benefits grow: How death can be outweighed by headaches' *Analysis* v58(2), pp152-158
- Pritchard, Duncan (2015) 'Risk' *Metaphilosophy* v46(3), pp436–461
- Rawls, John (1971) *A Theory of Justice* (Cambridge, MA: Harvard University Press)
- Russell, Jeffrey Sanford and Yoaav Isaacs (2021) 'Infinite prospects' *Philosophy and Phenomenological Research* v103(1), pp178-198
- Savage, Leonard (1954) *The Foundations of Statistics* (New York: Wiley)
- Smith, Martin (2010a) 'What else justification could be' *Noûs* v44(1), pp10-31
- Smith, Martin (2010b) 'A generalised lottery paradox for infinite probability spaces' *British Journal for the Philosophy of Science* v61(4), pp821-831
- Smith, Martin (2016) *Between Probability and Certainty: What Justifies Belief* (Oxford: Oxford University Press)
- Smith, Martin (2022a) 'The hardest paradox for closure' *Erkenntnis* v87(4), pp2003-2028
- Smith, Martin (2022b) 'Decision theory and de minimis risk' *Erkenntnis* online first: <https://link.springer.com/article/10.1007/s10670-022-00624-9>
- Smith, Nicholas (2014) 'Is evaluative compositionality a requirement of rationality?' *Mind* v123(490), pp457-502
- Spohn, Wolfgang (2012) *The Laws of Belief* (Oxford: Oxford University Press)
- Starkie, Thomas (1824) *A Practical Treatise of the Law of Evidence* (London: J. and W.T. Clarke), page references are for the *Seventh American Edition* (Philadelphia: T and J.W. Johnson, 1842)
- Steele, Katie and H. Orri Stefánsson (2016) 'Decision theory' in Zalta, E. ed *Stanford Encyclopedia of Philosophy* (Winter 2016 Edition)
URL = <https://plato.stanford.edu/archives/win2016/entries/decision-theory/>

Temkin, Larry (2001) 'Worries about continuity, expected utility theory, and practical reasoning' in D. Egonsson, D. Josefsson, J., Petersson, B. and Rønnow-Rasmussen, T. eds. *Exploring Practical Philosophy: From Action to Values* (Aldershot: Ashgate)

Thoma, Johanna (2019) 'Risk aversion and the long run' *Ethics* v129(2), pp230-253

Thomson, Judith Jarvis (1990) *The Realm of Rights* (Cambridge, MA: Harvard University Press)

Von Neumann, John, and Oskar Morgenstern (1947) *Theory of Games and Economic Behavior 2nd Edition* (Princeton: Princeton University Press)

Appendix: Stability of NDLEUT Decision Models

In section 7, I claimed that, although NDLEUT does not exhibit invariance under fine-graining, it has the following *stability* property: For any NDLEUT decision model, there is a more fine-grained model for which the normic de minimis expected utility of every action is unaffected by further fine-graining. In the appendix I make this claim precise and outline a proof.

We begin by setting things up more formally. In NDLEUT we have decision models of the form $\langle W, \Omega, \mathbf{A}, \mathbf{O}, \text{Pr}, \text{ab}, u_1, \dots, u_n, t \rangle$. W is a set of possible worlds and $\Omega \subseteq \wp(W)$ is a Boolean σ -algebra of propositions. \mathbf{A} and $\mathbf{O} \subseteq \Omega$ are partitions of W – sets of pairwise exclusive and jointly exhaustive nonempty subsets of W – which represent the set of actions and of outcomes respectively. If $\Pi \subseteq \Omega$ is a set of propositions, let $\text{cl}(\Pi)$ be the closure of Π under disjunction. Since \mathbf{O} is a refinement of \mathbf{A} we have it that $\text{cl}(\mathbf{A}) \subseteq \text{cl}(\mathbf{O})$ (with $\text{cl}(\mathbf{A}) = \text{cl}(\mathbf{O})$ in the case of a decision under certainty). Pr is a probability function taking the members of Ω into the set of real numbers in the unit interval and conforming to the axioms P1, P2 and P3, and ab is an abnormality function taking the members of Ω into the set of nonnegative integers plus ∞ and conforming to the axioms R1, R2 and R3. u_1, \dots, u_n are utility functions taking the members of \mathbf{O} into the set of positive and negative real numbers. Finally, t is some nonnegative integer, or ∞ .

Say that $\langle W, \Omega, \mathbf{A}, \mathbf{O}', \text{Pr}, \text{ab}, u_1', \dots, u_n', t \rangle$ is a *fine-graining* of $\langle W, \Omega, \mathbf{A}, \mathbf{O}, \text{Pr}, \text{ab}, u_1, \dots, u_n, t \rangle$ just in case $\text{cl}(\mathbf{O}) \subseteq \text{cl}(\mathbf{O}')$ and for all i , $1 \leq i \leq n$, u_i and u_i' agree for all outcomes in $\mathbf{O} \cap \mathbf{O}'$. Say that $\langle W, \Omega, \mathbf{A}, \mathbf{O}, \text{Pr}, \text{ab}, u_1, \dots, u_n, t \rangle$ is *normically calibrated* just in case for every outcome $O \in \mathbf{O}$ if X is a nonempty proposition in Ω such that $X \subseteq O$ then $\text{ab}(X) = \text{ab}(O)$. Less formally, $\langle W, \Omega, \mathbf{A}, \mathbf{O}, \text{Pr}, \text{ab}, u_1, \dots, u_n, t \rangle$ is normically calibrated just in case no outcome in \mathbf{O} 'cuts across' normalcy ranks – no outcome can be divided into more fine-grained outcomes that differ in terms of their normalcy.

Any NDLEUT decision model $\langle W, \Omega, \mathbf{A}, \mathbf{O}, \text{Pr}, \text{ab}, u_1, \dots, u_n, t \rangle$ has a fine-graining that is normically calibrated. *Proof* For each normalcy rank r , let $N_{\geq r} = \vee\{X \in \Omega \mid \text{ab}(X) \geq r\}$. $N_{\geq r}$ is the disjunction of all propositions that have an abnormality rank of at least r , and may be thought of as the set of worlds that have an abnormality of at least r . Since Ω is a σ -algebra (and closed under countable disjunction), $N_{\geq r}$ will be included in Ω , for each normalcy rank r . If there are no nonempty propositions in Ω that are assigned an abnormality rank of at least r then $N_{\geq r} = \emptyset$. By R3, $\text{ab}(N_{\geq r}) \geq r$.

For each r , let $N_r = N_{r \geq} \wedge \sim N_{r+1 \geq}$ (or $N_r = N_{r \geq} \wedge \sim N_{r >}$). N_r might be thought of as the set of worlds that have an abnormality of *exactly* r .²⁸ If there are no propositions in Ω that are assigned an abnormality rank of r then $N_r = \emptyset$. Since $N_r \subseteq N_{r \geq}$ it follows, by R3, that $\text{ab}(N_r) \geq r$. If $\text{ab}(N_r) > r$ then, by the definition of $N_{r+1 \geq}$, $N_r \subseteq N_{r+1 \geq}$. But, since $N_r \subseteq \sim N_{r+1 \geq}$ it follows that either $N_r = \emptyset$ or $\text{ab}(N_r) = r$. Let \mathbf{N} be the set of all nonempty propositions so defined.

Given a set of outcomes \mathbf{O} , let $\mathbf{O}^+ = \{O \wedge N_r \mid O \in \mathbf{O}, N_r \in \mathbf{N}, O \wedge N_r \neq \emptyset\}$. Since $O \wedge N_r \subseteq N_r$ it follows, from R3, that $\text{ab}(O \wedge N_r) \geq r$. If $\text{ab}(O \wedge N_r) > r$ then, by the definition of $N_{r+1 \geq}$, $O \wedge N_r \subseteq N_{r+1 \geq}$. But, since $O \wedge N_r \subseteq N_r \subseteq \sim N_{r+1 \geq}$, it follows that $\text{ab}(O \wedge N_r) = \text{ab}(N_r) = r$. Less formally, if an outcome O in \mathbf{O} crosses several normalcy ranks then, in \mathbf{O}^+ , O is divided according to these ranks. If, say, some of the worlds at which O is true have abnormality 1, some have abnormality 2 and some have abnormality 3 then, instead of O , \mathbf{O}^+ will contain three outcomes: O -in-the-abnormality-1-way (which would itself have an abnormality of 1), O -in-the-abnormality-2-way (which would itself have an abnormality of 2) and O -in-the-abnormality-3-way (which would itself have an abnormality of 3).

From the definition of \mathbf{O}^+ it follows that $\text{cl}(\mathbf{O}) \subseteq \text{cl}(\mathbf{O}^+)$ in which case if, for each i , $1 \leq i \leq n$, u_i is a utility function defined on \mathbf{O} and u_{i+} is a corresponding utility function defined on \mathbf{O}^+ such that u_i and u_{i+} agree for any outcomes in $\mathbf{O} \cap \mathbf{O}^+$ then $\langle W, \Omega, \mathbf{A}, \mathbf{O}^+, \text{Pr}, \text{ab}, u_{1+}, \dots, u_{n+}, t \rangle$ will meet the conditions for a fine-graining of $\langle W, \Omega, \mathbf{A}, \mathbf{O}, \text{Pr}, \text{ab}, u_1, \dots, u_n, t \rangle$. For any outcome O^+ in \mathbf{O}^+ , O^+ will be equal to $O \wedge N_r$ for some $O \in \mathbf{O}$ and $N_r \in \mathbf{N}$. Consider a nonempty proposition $X \subseteq O \wedge N_r$. Since $X \subseteq O \wedge N_r$ it follows, from R3, that $\text{ab}(X) \geq r$. If $\text{ab}(X) > r$ then, by the definition of $N_{r+1 \geq}$, $X \subseteq N_{r+1 \geq}$. But, since $X \subseteq O \wedge N_r \subseteq \sim N_{r+1 \geq}$ it follows that $\text{ab}(X) = \text{ab}(O \wedge N_r) = r$. It follows from this that $\langle W, \Omega, \mathbf{A}, \mathbf{O}^+, \text{Pr}, \text{ab}, u_{1+}, \dots, u_{n+}, t \rangle$ is normically calibrated. **QED**

Suppose we have a normically calibrated model $\langle W, \Omega, \mathbf{A}, \mathbf{O}, \text{Pr}, \text{ab}, u_1, \dots, u_n, t \rangle$ and a fine-graining $\langle W, \Omega, \mathbf{A}, \mathbf{O}', \text{Pr}, \text{ab}, u_1', \dots, u_n', t \rangle$ that meets the expected utility constraint. For any $A \in \mathbf{A}$, and any i , $1 \leq i \leq n$, $\text{NDEU}_i(A) = \text{NDEU}_i'(A)$. *Proof* Define a function f from the elements of \mathbf{O} to sets of elements of \mathbf{O}' such that, for each $O \in \mathbf{O}$, $f(O) = \{O_x \in \mathbf{O}' \mid O_x \subseteq O\}$. The function f , in effect, maps each outcome in \mathbf{O} to the set of sub-outcomes in \mathbf{O}' into which it is divided (and if an outcome is a member of both \mathbf{O} and \mathbf{O}' , f will map it to its own singleton). Note that, for any $O \in \mathbf{O}$, $O = \bigvee f(O)$ and, since outcomes are exclusive, f is a one-to-one correspondence between the elements of \mathbf{O} and a partition of the elements of \mathbf{O}' . Given normic calibration, for every $O \in \mathbf{O}$ and every $O_x \in f(O)$, $\text{ab}(O) = \text{ab}(O_x)$ from which it follows that $\bigwedge \{\sim O \mid O \in \mathbf{O} \wedge \text{ab}(O \mid \alpha(O)) > t\} = \bigwedge \{\sim O \mid O \in \mathbf{O}' \wedge \text{ab}(O \mid \alpha(O)) > t\}$ and $\psi = \psi'$. By the expected utility constraint, for any $O \in \mathbf{O}$, $u_i(O) = \sum_{O_x \in f(O)} \text{Pr}(O_x \mid O) \cdot u_i'(O_x)$.

Assume, for reductio, that there is an $A \in \mathbf{A}$ and an i , $1 \leq i \leq n$ such that $\text{NDEU}_i(A) \neq \text{NDEU}_i'(A)$. It follows that $\sum_{O \in \mathbf{O}} \text{Pr}_\psi(O \mid A) \cdot u_i(O) \neq \sum_{O \in \mathbf{O}'} \text{Pr}_{\psi'}(O \mid A) \cdot u_i'(O)$. By the definition of f ,

²⁸ If Ω contains the singleton of each world in W – that is, if $\{w\} \in \Omega$ for each $w \in W$ – then it will be the case that $N_{r \geq} = \{w \in W \mid \text{ab}(\{w\}) \geq r\}$ and $N_r = \{w \in W \mid \text{ab}(\{w\}) = r\}$. If Ω does not contain the singleton of each world in W – and this constraint is not demanded here – then the possible world descriptions of $N_{r \geq}$ and N_r given in the main text won't be reflected in the formalism (but can still serve as informal heuristics).

$\sum_{O \in \mathcal{O}} \Pr_{\psi'}(O | A) \cdot u_i'(O) = \sum_{O \in \mathcal{O}} \sum_{O_x \in f(O)} \Pr_{\psi'}(O_x | A) \cdot u_i'(O_x)$. But if $\sum_{O \in \mathcal{O}} \Pr_{\psi}(O | A) \cdot u_i(O) \neq \sum_{O \in \mathcal{O}} \sum_{O_x \in f(O)} \Pr_{\psi'}(O_x | A) \cdot u_i'(O_x)$ then there must be some $O^* \in \mathcal{O}$ such that $\Pr_{\psi}(O^* | A) \cdot u_i(O^*) \neq \sum_{O_x \in f(O^*)} \Pr_{\psi'}(O_x | A) \cdot u_i'(O_x)$. Suppose that $\text{ab}(O^* | A) > t$. It follows, by normic calibration, that for every $O_x \in f(O^*)$, $\text{ab}(O_x | A) > t$ in which case $\Pr_{\psi}(O^* | A) = 0$ and $\Pr_{\psi'}(O_x | A) = 0$ for every $O_x \in f(O^*)$. In this case $\Pr_{\psi}(O^* | A) \cdot u_i(O^*) = \sum_{O_x \in f(O^*)} \Pr_{\psi'}(O_x | A) \cdot u_i'(O_x) = 0$ contrary to assumption.

Suppose instead that $\text{ab}(O^* | A) \leq t$. Given normic calibration, it follows that for every $O_x \in f(O^*)$, $\text{ab}(O_x | A) \leq t$. It follows further that O^* entails ψ in which case, for any proposition $X \in \Omega$, $\Pr(X | O^*) = \Pr_{\psi}(X | O^*)$. Recall that, given normic calibration, $\psi = \psi'$. Since $u_i(O^*) = \sum_{O_x \in f(O^*)} \Pr(O_x | O^*) \cdot u_i'(O_x)$ we have it that:

$$\begin{aligned}
\Pr_{\psi}(O^* | A) \cdot u_i(O^*) &= \Pr_{\psi}(O^* | A) \cdot \sum_{O_x \in f(O^*)} \Pr(O_x | O^*) \cdot u_i'(O_x) \\
&= \Pr_{\psi}(O^* | A) \cdot \sum_{O_x \in f(O^*)} \Pr_{\psi}(O_x | O^*) \cdot u_i'(O_x) \\
&= \sum_{O_x \in f(O^*)} \Pr_{\psi}(O^* | A) \cdot \Pr_{\psi}(O_x | O^*) \cdot u_i'(O_x) \\
&= \sum_{O_x \in f(O^*)} (\Pr_{\psi}(O^*) / \Pr_{\psi}(A)) \cdot (\Pr_{\psi}(O_x) / \Pr_{\psi}(O^*)) \cdot u_i'(O_x) \\
&= \sum_{O_x \in f(O^*)} \Pr_{\psi}(O_x) / \Pr_{\psi}(A) \cdot u_i'(O_x) \\
&= \sum_{O_x \in f(O^*)} \Pr_{\psi}(O_x \wedge A) / \Pr_{\psi}(A) \cdot u_i'(O_x) \\
&= \sum_{O_x \in f(O^*)} \Pr_{\psi}(O_x | A) \cdot u_i'(O_x) \\
&= \sum_{O_x \in f(O^*)} \Pr_{\psi'}(O_x | A) \cdot u_i'(O_x)
\end{aligned}$$

Once again we have it that $\Pr_{\psi}(O^* | A) \cdot u_i(O^*) = \sum_{O_x \in f(O^*)} \Pr_{\psi'}(O_x | A) \cdot u_i'(O_x)$ contrary to assumption. It follows that, for every $A \in \mathbf{A}$ $\sum_{O \in \mathcal{O}} \Pr_{\psi}(O | A) \cdot u_i(O) = \sum_{O \in \mathcal{O}} \sum_{O_x \in f(O)} \Pr_{\psi'}(O_x | A) \cdot u_i'(O_x) = \sum_{O \in \mathcal{O}} \Pr_{\psi'}(O | A) \cdot u_i'(O)$ and $\text{NDEU}_i(A) = \text{NDEU}_i'(A)$. **QED**