Reverse-Engineering *Risk*

Lilith Mace and Angela O'Sullivan
Forthcoming in *Erkenntnis*, please cite published version when available.

Abstract

Three philosophical accounts of *risk* dominate the contemporary literature. On the probabilistic account, *risk* has to do with the probability of a disvaluable event obtaining; on the modal account, it has to do with the modal closeness of that event obtaining; on the normic account, it has to do with the normalcy of that event obtaining. The debate between these accounts has proceeded via counterexample-trading, with each account having some cases it explains better than others, and some cases that it cannot explain at all. In this article, we attempt to break the impasse between the three accounts of *risk* through a shift in methodology. We investigate the concept of *risk* via the method of conceptual reverse-engineering, whereby a theorist reconstructs the need that a concept serves for a group of agents in order to illuminate the shape of the concept: its intension and extension. We suggest that *risk* functions to meet our need to make decisions that reduce disvalue under conditions of uncertainty. Our project makes plausible that *risk* is a pluralist concept: meeting this need requires that *risk* takes different forms in different contexts. But our pluralism is principled: each of these different forms are part of one and the same concept, that has a 'core-to-periphery' structure, where the form the concept takes in typical cases (at its 'core') explains the form it takes in less typical cases (at its 'periphery'). We then apply our findings to *epistemic risk*, to resolve an ambiguity in how 'epistemic risk' is standardly understood.

Keywords: risk, epistemic risk, conceptual reverse-engineering, anti-risk epistemology

Reverse-Engineering *Risk*

> "Many of you here remember that when our Society for Risk Analysis was brand new, one of the first things it did was to establish a committee to define the word 'risk'. This committee labored for 4 years and then gave up, saying in its final report, that maybe it's better not to define risk. Let each author define it in his own way, only please each should explain clearly what way that is."

Stan Kaplan 1997, address at the 1996 Annual Meeting for the Society of Risk Analysis.

1. Introduction

The concept *risk* is central to many aspects of human life. In day-to-day life, we use *risk* to guide our decision-making. For example, if I am considering whether to get laser eye surgery, I will weigh the potential benefits of the surgery against the risk of it going wrong. *Risk* finds widespread use in industrial contexts, such as engineering, banking, security and waste-management – indeed, in any industry that deals with potential harm or loss of valuable resources. Within philosophy, ethicists debate the effects of *risk* on right action (Thomson 1983; Buchak 2017; Thoma 2019; Lee-Stronach 2019), while epistemologists develop theories of knowledge on which knowledge is incompatible with high levels of risk in the epistemic realm (Pritchard 2015, 2016; Navarro 2019, 2021).

Yet despite the significance of the concept *risk* in these varied contexts, it is far from clear how the concept should be understood. Some risk theorists have expressed pessimism that any cogent account of risk is possible, let alone forthcoming.[1] In recent years, philosophers have attempted to clarify the nature of *risk* in terms of notions supposedly better understood: evidential probability, modal closeness, and normalcy. The resulting accounts of *risk* – the probabilistic account; the modal account, proposed by Duncan Pritchard (2015); and the normic account, proposed by Philip Ebert, Martin Smith and Ian Durbach (2020) – generate incompatible risk-evaluations. As such, one cannot accept all three as correct descriptions of one monist concept *risk*.

In this paper, we illuminate *risk* through the method of conceptual reverse-engineering, whereby a theorist reconstructs the needs that a concept serves, to illuminate its 'shape': its intension and extension. We argue that *risk* serves its function by varying its content in different contexts: in some contexts, its content is as the probabilistic account has it; in others, it is as the modal account has it; in yet others, it is as the normic account has it. Our project thereby makes plausible that *risk* is a pluralistic concept, as suggested by Ebert, Smith and Durbach (2020); though our account of this pluralistic concept improves on that offered by Ebert, Smith and Durbach, in that it both explains why *risk* is pluralist, and how the different forms *risk* takes relate to each other: we argue that *risk* has a core-to-periphery structure (Fricker 2008), taking a different form in typical ('core') cases than it does in less typical ('peripheral') cases.

We then apply this picture to the epistemic realm, to resolve an ambiguity in recent epistemological literature on *epistemic risk*. The phrase 'epistemic risk' is used in two ways in the literature. On the first, 'epistemic risk' is used to talk about a variety of epistemically

[1] As well as the quote from Kaplan, the discussion on "Managing the undefinable" from the U.K. Government's National Cyber-Security Centre website (2018) takes for granted that no comprehensive definition of 'risk' is possible. There, an attempt is made to spin this as a valuable aspect of risk-management: the uncertainty keeps different speakers and authors on the alert for possible miscommunication.

disvaluable events: forming a false belief, failing to form a true belief, obtaining misleading evidence, failing to obtain good evidence, failing to know, and so on. We argue that this use of 'epistemic risk' picks out a concept *epistemic risk*, which has the same core-to-periphery structure as *risk*. But on the second, more popular way of using 'epistemic risk', it picks out only the risk of forming a false belief. We will explain why this second use of 'epistemic risk' is found more often in the literature than the first by appealing to the core-to-periphery structure of *epistemic risk*. We will argue that epistemologists working with the concept are interested in peripheral cases, and those cases tend to be such that the only relevant epistemic risk-event is the event of a subject's forming a false belief.

2. Three accounts of *risk*

The "standard" or "orthodox" account of *risk* (so-called by Pritchard 2015, p. 436; Bricker 2018, p. 200; Ebert, Smith and Durbach 2020, p. 432) is the probabilistic account. On the probabilistic account, risk-events are disvaluable events with a non-zero probability of occurring, given a body of evidence; high-risk events are disvaluable events with a high probability of occurring and low-risk events are disvaluable events with a low probability of occurring, with a continuum of riskiness between these extremes; and an event E1 is higher risk than an event E2 if the probability of E1's occurring is higher than the probability of E2's occurring. The probabilistic account of *risk* says, for example, that there is a very low risk that I will be killed by lightning strike this year, as there is a very low (but non-zero) probability that this event will occur, relative to my evidence: about one in 19 million, or 0.0000012 (Elsom 2001). In contrast, there is a high risk of dying when playing Russian roulette: just under 1 in 6, or 0.1666…[2].

Despite its orthodoxy, Pritchard argues that the probabilistic account of risk is fatally undermined by its inability to account for our intuitions regarding the following pair of cases (2015, p. 441, here lightly rephrased):

> **Bomb 1**. An evil scientist has hidden a bomb in a highly populated area. The bomb is rigged to detonate if a certain set of numbers comes up on the next national lottery draw. The odds of these numbers coming up is fourteen million to one. There is no way of disarming the bomb before it is set to detonate.
>
> **Bomb 2**. An evil scientist has hidden a bomb in a highly populated area. The bomb is rigged to detonate if a series of three highly unlikely events occur. First, the weakest horse in the Grand National must win the race by at least ten furlongs. Second, the worst team remaining in the FA Cup draw, Accrington Stanley, must beat the best team

[2] 'Just under' to account for the tiny chance that the gun will misfire, or that the player will miraculously survive getting shot.

> remaining, Manchester United, by at least ten goals. Finally, the King of England must spontaneously decide to speak a complete sentence in Polish in his next public speech. The odds of these three events occurring is fourteen million to one. There is no way of disarming the bomb before it is set to detonate.

The probability of the bomb detonating in each of Bomb 1 and Bomb 2 is, by stipulation, identical. Despite this, Pritchard holds that the situation in Bomb 1 is "clearly far more risky" than the situation in Bomb 2, since the bomb blast in Bomb 1 is, "something that *could very easily occur*" (2015, pp. 441-2). The probabilistic account cannot explain this divergence: as the probability of the bomb going off is equal between the cases, so is the risk. Hence Pritchard concludes that the probabilistic account is "fundamentally misguided" (436).

To replace the probabilistic account, Pritchard proposes a novel account of *risk*, which he calls the "modal account" (2015, p. 436). On the modal account, risk-events are disvaluable events that obtain in some possible world; high-risk events obtain in close possible worlds, where close possible worlds are worlds that are similar to the actual world (Lewis 1973); low-risk events obtain in distant possible worlds, where distant possible worlds are dissimilar to the actual world; and the risk of an event E1 is higher than that of an event E2 if the closest world in which E1 obtains is closer to the actual world than the closest world in which E2 obtains (Pritchard 2015, p. 447). Unlike the probabilistic account, the modal account does not relativise risk to a body of evidence: the level of risk involved in a given situation is determined solely by how the actual world is, and how much would have to change to get from the actual world to a world in which the risk-event obtains; whether any body of evidence suggests that the actual world is this way, or that so much would have to change to get from the actual world to the risk-event world, makes no difference to the level of risk in play. On the modal account, there is a low risk of me being killed by lightning strike if this isn't something that occurs in a close world; if, for example, I make sure to never be outside during a thunderstorm. In contrast, the risk of me being killed by lightning is high if I sit on the roof of a skyscraper, holding a metal antenna, during a thunderstorm. This is so even if the probability that I will get struck by lightning in this situation is not high.

Pritchard's account gets the intuitively correct[3] result in his bomb cases: that the risk of the bomb detonating in Bomb 1 is higher than in Bomb 2. For the closest world in which the bomb detonates in Bomb 1 is very close indeed: a few coloured balls need only fall in a certain configuration. But the closest world in which the detonation-triggering conditions obtain in Bomb 2 is not close at all. Given the way that the actual world is, it could not easily happen that

[3] At least according to Pritchard's intuitions. In a survey of non-philosophers conducted by Ebert, Smith and Durbach, it was found that although those surveyed tended to say they would *prefer* to be in the Bomb 2 scenario than the Bomb 1 scenario, they nevertheless tended to judge the two scenarios as equally risky (2020: 450-1).

the weakest horse in the Grand National wins the race by ten furlongs, that the worst team in the FA Cup beat the best team by ten goals, or that the King of England spontaneously chooses to speak a complete sentence of Polish in his next speech, let alone all three. As the closest world in which the bomb detonates in Bomb 1 is closer than the closest world in which the bomb detonates in Bomb 2, the scenario of Bomb 1 is riskier than that of Bomb 2.

However Pritchard's modal account also faces counterexamples. As Ebert, Smith and Durbach (2020) and Fratantonio (2021) note, it is a consequence of the modal account that any actually obtaining risk-event is maximally risky. For Pritchard has it that the closer is the closest world in which the risk-event obtains, the riskier is that event, and the actual world is maximally close (Ebert, Smith and Durbach 2020, p. 441). It is also a consequence of Pritchard's account that any actually obtaining risk-event is riskier than any risk-event that does not actually obtain; for however risky is a non-obtaining risk-event, it is less than maximally risky. But this is implausible. To illustrate, consider a modification of Ebert, Smith and Durbach's example, in which someone is about to drill into the wall of a West Australian house built in the 1970s, and is wondering about the risk of the wall's containing asbestos (2020, p. 441). On Pritchard's modal account, if the wall actually contains asbestos, then the risk of the wall's containing asbestos is maximal, while if the wall doesn't actually contain asbestos, then the risk is less than maximal. Imagine (and here is the modification of the original case) that the driller's neighbour is in the same situation: he too is about to drill into the wall of his 1970s-built house, and is wondering about the risk of the wall's containing asbestos. Suppose that both drillers have the same evidence for thinking there might be asbestos in their wall. But suppose that only the first neighbour's wall contains asbestos. Pritchard's modal account has it that the risk of the first neighbour's wall containing asbestos is higher than the risk of the second neighbour's wall containing asbestos, even though they both have the same evidence for thinking their wall might contain asbestos, and as such they ought to take exactly the same steps before drilling into the wall. This is an uneasy result: intuitively, the risk of both walls containing asbestos is the same.

One might wonder whether this example could be accommodated on the modal account by relativising the modal notion of *risk* to a body of evidence. An initial obstacle is that closeness ordering on worlds model the extent to which different worlds resemble the actual world, not the extent to which any (non-maximal) body of evidence suggests that they resemble the actual world (Newton, 2022). A body of evidence can suggest that a given world is close to the actual world, but it cannot make this so – unless the body of evidence is the maximal body of evidence, containing all and only the true propositions about our world. As such, making this change would mean that the modal theory no longer appeals to a closeness ordering on the actual world, but a similarity ordering relative to some set of other possible worlds, which may or may not include the actual world. This would represent a substantial departure from Pritchard's modal

theory.[4] Further, as we explain in §3.3, the modal account is explanatorily powerful in epistemic contexts precisely because it captures an objective, evidence-free ordering. For this reason, among others, Pritchard rejects the idea of relativising his modal notion of *risk* to a body of evidence. Instead, he accounts for the West Australian House case by positing a distinction between the "actual risk in play" in a given context, vs. what would be a "reasonable risk assessment" in that context, arguing that the latter, but not the former, is relativised to evidence (2022a, p. 290). Pritchard acknowledges that in cases where one's evidence is incomplete or misleading, the modal account has the consequence that what is reasonable to judge about the level of risk in a case will diverge from the actual level of risk in play. But he contends that this is "simply a consequence of the fact that this proposal treats risk as an objective feature of the world" (2022a, p. 289), and objective features of the world are in general such that our reasonable judgements about them are often mistaken, on account of being based on misleading evidence. The point remains, Pritchard insists, that when we are making judgements about risk, what we are trying to capture is the modal closeness of a negative event obtaining, and as such, the modal account of *risk* is what guides our risk judgements.

Yet Pritchard's distinction between actual risk and reasonable risk judgement does not salvage the usefulness of the modal account of *risk* when it comes to evaluating risk. For as Smith (2023) points out, the modal account collapses the difference between one's reasonable judgment about the *risk* of an event obtaining, and one's reasonable judgements about *whether that event in actual fact obtains*. Recall that the modal account has the dual consequences that, in the West Australian House case, if there is a low risk that the wall does not contain asbestos, then the wall does not contain asbestos; and if the wall does contain asbestos, then there is a maximal risk that the wall contains asbestos. What this means is that if one is not in a position to make the judgment that the wall does not contain asbestos, they are equally unable to make a judgment that *there is a low risk* that the wall contains asbestos. As Smith puts the point, when it comes to judging the level of risk according to the modal account, "risk effectively collapses into truth" (2023, p. 156): we can (reasonably) judge that an event is high risk iff we can judge that it actually obtains, or will obtain, and we can (reasonably) judge that an event is low risk iff we can reasonably judge that it doesn't, or won't, obtain.

In response to this problem, Ebert, Smith and Durbach (2020) propose yet another novel account of risk, on which what determines risk is not how close the worlds in which a risk-event obtains

[4] A theorist could develop a distinct theory of risk on which risk is determined by a similarity order on a set of worlds consistent with a subject's evidence. This would be a very different view to Pritchard's - in particular, it would not capture a notion of "risk as an objective phenomenon" that Pritchard cares about (2015, p. 440) - but it would handle the West Australian House case. It would take us too far beyond the bounds of our paper to develop such an account of risk ourselves. We note, however, that such an account would still face counterexamples: it would predict that the two bomb blasts are equally likely in Ebert, Smith and Durbach's third bomb case (2020, p. 446), which we discuss at the end of this section.

are, but how normal those worlds are. The notion of normalcy appealed to is that developed by Smith (2016) in terms of calling out for explanation. The obtaining of an event E is normal, in Smith's sense, if E's obtaining would not call out for special explanation, given a body of evidence (2016, p. 39). Whether something calls out for special explanation is not a fully subjective matter, in that P does not call out for special explanation iff some subject wants an explanation of P. Rather, once a body of evidence is fixed, whether P calls out for explanation is likewise fixed, whether or not any subject realises - or cares - that this is so. Suppose that I see what looks to me like a red mug on a black table. If this mug were not red, special explanation would be called for: perhaps the mug is bathed in red light, perhaps I am hallucinating. So the mug's not being red is abnormal, in Smith's sense. But this doesn't require that *I*, or anyone else, *wants* an explanation for how the mug could turn out not to be red, despite appearing red to me. Rather, given the body of evidence I have, it simply is the case that the mug's not being red would call out for special explanation. Further, the mug's being red, given my evidence, would *not* call out for special explanation; as such, the mug's being red is normal. This, similarly, is not because I (or anyone else) doesn't *want* an explanation for this fact. Given my evidence, the fact simply does not call out for explanation, whether or not anyone wants an explanation.

Possible worlds can be ordered in terms of their normalcy (Smith 2016, p. 42). The most normal worlds are those worlds whose obtaining would call out for no explanation; worlds become less normal as their obtaining would call out for more explanation. Given this picture of normalcy, Ebert, Smith and Durbach offer their "normic account" of *risk*, according to which risk-events are disvaluable events that obtain in some possible world; high-risk events obtain in normal worlds; low-risk events obtain in abnormal worlds, where an abnormal world is a world whose obtaining would call out for special explanation, given a body of evidence; and an event E1 is higher risk than an event E2 if the most normal world in which E1 obtains is more normal than the most normal world in which E2 obtains (2020, p. 443-4).

To see how the normic account differs from the modal account, consider some examples. On the normic account, there is a low risk, relative to my evidence, that my partner has missed her train home. The obtaining of this event would call out for explanation, given my evidence, which includes such facts as that the train service is reliable, and that my partner left the office on time. This is so even if she has, in fact, missed her train home. The modal account, in contrast, would say that in this case there is a high risk that my partner has missed her train home, because this event obtains in a maximally close world: the actual world. On the normic account, there is a high risk, relative to my evidence, that I will have a stomach ache later tonight, because I know I am lactose-intolerant, but I nevertheless had a cheese sandwich for lunch. This is so even if, unbeknownst to me, the 'cheese' was vegan cheese, so did not contain lactose; in this case, the modal account issues the verdict that there is a low risk of my having a stomach ache later tonight.

The debate between the probabilistic, modal, and normic accounts of *risk* has proceeded largely by trading counter-examples. Currently, each of the three accounts has some cases for which it issues intuitively correct verdicts, and some cases for which it does not. Though both the modal and normic accounts have the advantage over the probabilistic account of issuing the intuitively correct verdict in Pritchard's original bomb cases, the probabilistic account has an advantage over the modal and normic accounts in a third bomb case devised by Ebert, Smith and Durbach. In this case, the bomb will go off if a certain set of numbers comes up in the next lottery, but in one scenario the probability of these numbers coming up is one in fourteen million, and in the other it is one in one billion (2020, p. 446). The probabilistic account generates the intuitively correct verdict that there is a much higher risk of the bomb's detonating in the first scenario than in the second. But the modal and normic accounts generate the result that the bomb blast is equally risky in both scenarios, because in both cases it occurs in a world that is equally close and equally normal, respectively. Further counterexamples to the modal account have been offered by Adam Bricker (2018, p. 203) and Ebert, Smith and Durbach (2020, p. 443), and counterexamples to the normic account by Marvin Backes (2018, p. 2884).

The debate has reached an impasse, with no clear victor. Each account generates counterintuitive results in at least one test case. We suggest trying a new approach. Instead of assuming that we have intuitive access to the extension of the concept *risk*, such that we can test each proposed theory of *risk* by how well it captures this intuitive extension, we will begin our investigation into *risk* by asking what the concept does for us: what having *risk* in our conceptual repertoire enables us to do that we couldn't do, or couldn't do as easily, if we lacked the concept. We will develop and motivate a hypothesis about the function of *risk*, then theorise about what the concept must be like to successfully fulfil this function. This will, we hope, breathe new life into the stagnating debate on the nature of *risk*.

3. Conceptually reverse-engineering *risk*

We investigate *risk* using the method of conceptual reverse-engineering, whereby a theorist sheds light on the 'shape' of a concept – its intension and extension – by reconstructing the practical needs that this concept meets for some group of agents (Queloz 2021, p. 53). The motivating idea behind the method is that many of our concepts, *risk* included, emerge and remain in circulation because they serve particular purposes: they enable a group of agents to achieve something that they could not, or could not so easily, achieve without that concept. If we are interested in investigating a concept's intension and extension, we may proceed by first identifying what function the target concept fulfils, and then "reverse engineer" (Queloz 2021, p. 16) the concept by asking what intension and extension the concept would require in order to fulfil the posited function.

Conceptual reverse-engineering can take many different forms[5]. One form, which we adopt in this paper, begins by offering a "plausible hypothesis" (Craig 1990, p. 2) about the function of the target concept: one theorises about which of our needs are fulfilled by having the target concept as part of our conceptual repertoire. Second, one identifies or constructs a typical case in which this need is present. The case is 'typical' not in the sense of being a frequently encountered scenario, but in being representative of situations of the relevant type, namely, ones which feature the hypothesised need. Third, the conceptual reverse-engineer asks: what will a concept that serves this need in the typical case look like (Craig 1990, p. 3)? What will its intension and extension be? A concept with this intension and extension is then posited as the concept that serves the need for those agents in the typical case. Finally, the conceptual reverse-engineer compares the concept that emerges in the typical case to the "intuitive" concept of interest (Craig 1990, p. 2): the concept that we actually use, with the intension and extension that are suggested by our use. If the emergent concept is recognisably similar to the intuitive concept, then the typical case can be understood as capturing "the most simple and basic form of the extant practice" that we have with the concept (Fricker 2016, p. 165). The emergent concept will correspondingly be understood as constituting the "core" of the concept (Fricker 2008, p. 40), which may be elaborated in different ways to meet various local needs, thus changing the content of the concept; but these elaborations are to be understood as elaborations of the basic, 'core' form of the concept (Queloz 2021, p. 27).

Before proceeding, note that we don't see conceptual reverse-engineering as a replacement for the traditional method of delineating a concept through considering counterexamples. Rather, conceptual reverse-engineering is an additional tool in the philosopher's methodological toolbox. Counterexample-trading can show that a proposed intension for a concept should be rejected, because the extension it demarcates conflicts with the concept's intuitive extension. Conceptual reverse-engineering can show that, if a concept functions in a certain way, it will have a particular intension and extension. Both of these methods share the same aim: accurately describing a concept's intension and extension. A philosopher can thus make use of both methods in a project with this aim.

Furthermore, we don't think that any and all philosophical inquiries necessitate the method of conceptual reverse-engineering. However, we do think conceptual reverse-engineering is particularly well-suited for inquiring into the nature of *risk*. First, the method of counterexample-trading has landed the *risk* debate in somewhat of a stalemate. Second, for other concepts that philosophers have tried to reverse-engineer, such as *knowledge*, a lot of work must be done to motivate the idea that the concept is functional in the first place (see for example Hannon 2019, ch. 2; Queloz ch. 3, §2). In contrast, *risk* is a concept that wears its functionality

[5] Some of the forms that it can take are: genealogies (e.g. Hume 1739, Craig 1990, and Williams 2002); models (e.g. Wittgenstein 1953, Hannon 2019, Queloz 2021); paradigm-based explanations (e.g. Fricker 2016).

on its sleeve: it is clearly useful for creatures like us to think and talk in terms of *risk*. Thus, even if we cannot do so for concepts that are less obviously functional, we should expect to be able to illuminate *risk* by reflecting on its purpose.

3.1. The function of risk

Conceptual reverse-engineering begins with a hypothesis about the function of the target concept: with a hypothesis about what having the concept enables some group of agents to do that they couldn't do, or couldn't as easily or as efficiently do, without the concept (Gardiner 2015, p. 31; Hannon 2019, p. 12; Thomasson 2020, p. 445). What does having *risk* in our conceptual repertoire enable us to do that we couldn't (easily) do if we lacked this concept? One way to answer this is to imagine people much like us – who have the same biological needs for food, water, shelter; who live socially and use language; and so on – but who lack the concept *risk*, and think about what needs of theirs would go unfulfilled (Craig 1990, Williams 2002).

These people are able to reason about what to do when some course of action is certain, or highly likely, to result in significant harm to oneself or to others, as these situations fall under the concept *danger*. If one such person is fording a river and sees a crocodile approaching, then she takes herself to be in danger and does everything in her power to remove herself from the dangerous situation. These people are also able to reason about situations in which some course of action is certain not to have any negative outcome. If a person knows that this section of the river is devoid of crocodiles because it is too saline for them, then she doesn't need the concept *risk* to decide whether to ford the river at this point. However, these people will also sometimes have to make decision about crossing the river that do not turn on whether the crossing would be dangerous. Say that they need to decide whether to travel upstream or downstream to ford the river: upstream is three times the distance, but depending on the speed at which the snow from the mountains several miles away is melting, downstream may have become too deep to cross. What this agent needs is a concept of risk to apply in weighing the potential disvalue of travelling three times the distance to the upstream ford, compared to the potential disvalue of having to turn back if the downstream ford is flooded. Without the concept *risk*, the agent cannot efficiently or effectively reason about how to cross the river. The concept of *danger* is not helpful: in this situation, there is no danger; and reasoning as if either the long journey or the flooded ford were dangerous would not reduce disvalue, as it would mean potentially avoiding crossing altogether, which let us assume would itself be disvaluable for the agent. So the concept of *danger* here is not action-guiding; but there is a decision to be made, with varying amounts of disvalue depending on what the agent chooses.

The concept *risk* would enable these people to reason about potential disvalue that is not guaranteed to occur, but whose occurrence is "realistically possible" (Pritchard 2015, p. 429);

something which "might actually", as opposed to "merely might", occur (Grimm 2015, p. 132; see also Blome-Tillmann on those possibilities that are "live options" 2009, p 247). Whether our river-crossing subject should go upstream or downstream hinges on the extent to which the downstream ford may well be flooded and on the severity of the disvalue should she have to turn back, relative to the severity of the disvalue of travelling for three times as long to the upstream ford. In the absence of a concept with which to conceptualise these two inter-related dimensions of the situation – the extent to which disvalue may occur, and the severity of the disvalue – agents cannot (easily, efficiently) compare different courses of action with the goal of reducing disvalue[6]. Thus we hypothesise: *risk* functions to guide decision-making so as to reduce disvalue under conditions of uncertainty. The concept *risk* is therefore at its most functional when negative outcomes are not guaranteed to arise, but in which they might – might actually – do so.

That *risk* functions in this way is supported by reflection on everyday cases. You must decide whether to catch the bus or take a taxi, weighing up whether disvalue might well obtain in each scenario, and how severe this disvalue would be. If you must get to the airport on time to catch a flight, the guaranteed disvalue of paying an extortionate taxi fare might be worth the trade-off of avoiding the potential disvalue of the bus being late, which might well happen. If you're on your way home with no evening plans, this trade-off likely won't be worth it. In any case, you can appeal to *risk* to help you make your decision: when it comes to catching the flight, getting the bus is too risky; when it comes to getting home after work, getting the bus is not risky at all. (Note further that the concept *danger* will not aid your reasoning here. You're not in danger in any of these cases.)

3.2. The core of risk

Thus we have a plausible hypothesis from which to begin our conceptual reverse-engineering project: *risk* functions to guide decision-making so as to reduce disvalue under conditions of uncertainty. We will now construct a typical case containing the need for a concept that functions in this way. Imagine an agent considering whether to run a marathon. Running a marathon could

[6] One might object that our agent in this situation does not need a new concept with which to navigate uncertainty, but could just as well follow a series of default rules or procedures, e.g. "never ford the river downstream in summer or after heavy rain" combined with "never ford the river upstream in winter". We hold that such agents would still have a need for the concept of risk, for two reasons: first, given how pervasive conditions of uncertainty are (and how variable potential disvalue is), such agents would often need to make decisions under novel conditions of uncertainty for which they possess no pre-existing default rule (e.g. it is an unprecedentedly warm winter, and so the ice on the mountain may well have melted and flooded the downstream ford – does the policy about never fording the upstream ford in winter still apply?). Second, whether or not agents could in principle navigate uncertainty so as to reduce disvalue without the concept of *risk* is not the only consideration. The conceptual engineer also wants to know if agents could do so "as effectively or efficiently" (Thomasson 2020, p. 448) without the target concept. In this case, memorising and following a vast amount of predetermined context-specific rules would be significantly less effective and efficient than having a single concept which enables us to reason about, and compare, levels of potential disvalue.

result in various kinds of disvalue: pulled muscles, damaged kidneys, increased cortisol levels and even death from cardiac arrest. Suppose that the kind of disvalue with which the agent is most concerned is death from cardiac arrest. For all she knows, she is very fit and healthy. In particular, she's never been diagnosed with any heart conditions. However, most cardiac arrest marathon deaths are due to underlying coronary artery conditions, such as hypertrophic cardiomyopathy, which typically go undetected. What must *risk* be like to serve its function in this case? In particular, must it be like the concept of *risk* demarcated by any of the three accounts of *risk*? We will henceforth call the concepts demarcated by the modal, probabilistic and normic accounts '*modal risk*', '*probabilistic risk*' and '*normic risk*', respectively.

The modal account of risk is entirely unhelpful for guiding the agent's decision-making in this case. What the modal account tells our agent is that, *if* she has an underlying heart condition, *then* she's at high risk of death and should not run the marathon; whereas *if* she has no underlying heart condition, *then* she's at low risk of death and can run the marathon. But whether she has an underlying heart condition or not is precisely what our agent doesn't know: given that underlying heart conditions present with no symptoms, both options are live from her perspective. Then she cannot determine whether she is at high or low risk of cardiac arrest without knowing which medical condition she has. Ebert, Smith and Durbach (2020) and Fratantonio (2021) point out that, on the modal account, one cannot make a judgement about risk without taking a stance on whether the relevant risk-event obtains. We add to this that one cannot make a risk-judgement without taking a stance on what one's situation is like in general. But typical cases in which an agent needs to appeal to *risk* are cases in which she is ignorant of many details of her situation. Furthermore, as discussed in §2, the problem is not solved by relativising the risk to the marathon runner's evidence. Because the modal account collapses the difference between reasonable judgements about the risk of an event obtaining and reasonable judgements about whether the event in actual fact obtains, then the would-be marathon runner can only (reasonably) judge whether she is at low risk of suffering from cardiac arrest if she can (reasonably) judge whether a fatal cardiac arrest will occur. But of course, if she can reasonably judge whether a fatal cardiac arrest will occur, then she has no need of the concept of risk to guide her decision making.

Thus the notion of *risk* demarcated by the modal account cannot serve the hypothesised function of *risk* in this case: it cannot guide our agent in her decision-making.

Does the probabilistic notion of *risk* do any better? Somewhat. On the probabilistic account of *risk*, it is at least possible for the agent to determine the risk of her dying from cardiac arrest during the marathon, given that she has experienced no symptoms. The would-be-marathon-runner might start from data on how many people die from cardiac arrest during or immediately after running marathons. A recent medical review found that there are between 0.6 and 1.9 sudden cardiac deaths during or immediately after marathons per 100,000

runners (Waite et al. 2016); for ease of explication, let's say the number is one in 100,000. Then the probability of suffering a cardiac arrest death, on her evidence, is one in 100,000. This is pretty low, so cardiac arrest death during or immediately after the marathon is low-risk, on the probabilistic account.

However, this number is not, by itself, particularly informative to the individual decision-maker. This is for two reasons. The first is that making an accurate probabilistic calculation that is sufficiently relevant to a particular person is a very complicated matter. For one thing, population-wide probabilities can often differ significantly from probabilities concerning demographics within a population. For example, one review study (Kim at al 2012) found that men are significantly more likely than women to suffer from cardiac arrest during or immediately after a marathon (0.90 / 100,000 runners compared to 0.16 /100,000 respectively); and of those who do experience cardiac arrests, previous marathon completion correlates positively with survival (survivors had an average of 3.5 previously completed marathons, compared to non-survivors' 1.5). Further, one case report on cardiac arrest during marathons found that an "accurate determination of the incidence of the phenomenon is very difficult to achieve, because of the extreme differences in age, sex, race, athletes and non-athletes" (Ghio et al, 2012, p. 130). So in order to make a probability calculation that is sufficiently informative for oneself in particular, one needs statistics not at the population level, but for more specific demographics of which they are a member.

But even once one has more specific information from which they can calculate the probability of a given risk event obtaining, *actually calculating* that probability is a complicated matter. Our would-be marathon runner is trying to work out the risk of her dying of cardiac arrest if she were to run the marathon. She knows that she has no symptoms of an underlying heart condition. So in order to calculate this risk, she needs to work out the probability of someone sufficiently like her dying of cardiac arrest during the marathon, conditional on not having any symptoms of underlying heart conditions. Let 'S' be the proposition that such a person is symptomless, and let 'C' be the proposition that such a person dies of cardiac arrest during the marathon. The agent needs to know the prior probability of S, the prior probability of C, and the probability of S given C. From this, she can calculate the probability of a given person sufficiently like her dying of cardiac arrest, conditional on having no symptoms, using Bayes' Theorem: $Pr(C|S) = Pr(S|C).Pr(C)/Pr(S)$.

We have deliberately not plugged in specific numerical values for our variables, here. This was not (only) because these probabilities are themselves not straightforward to determine, but because the problem that we are getting at does not turn on the actual output of any given probability calculus. Our objection is not that the probabilistic account issues risk judgements

that are implausible. Rather, it is that it is very difficult for a given agent to determine the level of probabilistic risk involved in her situation, so much so that *probabilistic risk* is not plausibly the notion of *risk* underpinning everyday decision-making. In order to work out the level of probabilistic risk in a given situation, a subject must first know the prior probabilities of many propositions. Some of these will be more difficult to find out than others (in our case, how is our would-be marathon runner to find out how many of the runners who died of cardiac arrest did not have any symptoms, i.e. Pr(C|S)?). Second, she must be able to perform complicated calculations (for example, Bayesian conditionalisations) that most people cannot do; or at least, cannot do as quickly and efficiently as using the notion of *risk* in everyday decisions would demand. The probabilistic account of risk thus yields a *risk* concept that is usefully action-guiding for some mathematicians and fewer philosophers, and hardly anyone else.

The second reason that the probabilistic account of *risk* is not particularly useful for guiding behaviour in the core case is that it issues a notion of *risk* that is impersonal. An agent can use the probabilistic notion of *risk* to determine that X% of people like her, in situations like hers, would experience the risk event. But this doesn't tell her anything about whether *she in particular* would be in the unlucky X%. In our case, it may be that the probability of any given person sufficiently like our would-be marathon runner having hypertrophic cardiomyopathy might be low; in that case, the probabilistic notion of risk tells our runner that the risk of dying of cardiac arrest during the marathon is very low (as not all people with hypertrophic cardiomyopathy would go into cardiac arrest during the marathon, and not all those who go into cardiac arrest would die). But for all our agent knows, she could be one of the people who do suffer from hypertrophic cardiomyopathy, rather than one of the people with no underlying heart condition. In that case, it would be no comfort at all to know that *most* people like her do not have a heart condition. Another way of putting this point is in terms of perspectives. From an outsider's perspective, someone can reason as follows: it is unlikely that a given person with no symptoms has hypertrophic cardiomyopathy, and thus even less likely that a given symptomless person would suffer a fatal cardiac arrest during a marathon; so there is a low risk that a given symptomless person would suffer a fatal cardiac arrest during the marathon; so there is a low risk that *this* symptomless person would suffer a fatal cardiac arrest during the marathon. But from the agent's own perspective, it is more difficult to make the jump from 'there is a low risk that a given symptomless person would suffer a fatal cardiac arrest during the marathon' to 'there is a low risk that *I* would suffer a fatal cardiac arrest during the marathon' - after all, to quote Lina Lamont, "I *ain't* people*"*. As such, the usefulness of the probabilistic notion of *risk* in guiding our agent's behaviour with the aim of reducing disvalue for her in particular is limited.[7] Or at

[7] One might worry that we have been uncharitably literal in characterising what the probabilistic account of risk demands of agents seeking to use *probabilistic risk* in their daily decision making. After all, the objection might go, one need not calculate the exact probability given one's own demographic – one need only calculate whether the probability is sufficiently low, irrespective of one's demographic. In this case, one might think that the probability of

least, it is limited for "creatures like us" (Queloz 2021, p.1) – that is to say, creatures with limited access to relevant statistics, and limited insight into which statistics would best capture the probability for us, given our particular characteristics.

In contrast, *normic risk* can usefully guide decision-making. If the would-be marathon runner has no evidence that she has an underlying heart condition, then her dying as a result of running the marathon would cry out for explanation (perhaps the explanation would be precisely that she has an undetected underlying heart condition). *Her death in particular* is thus low-risk.[8] She can determine this by reflection on her evidence. This requires no knowledge beyond her ken, no complex calculations, and no reasoning about how population-wide probabilities relate to her as an individual. As such, the subject has all the information needed to make a reasonable judgement of normic risk, and can do so quickly, with relative ease. Thus *normic risk* fulfils the function of *risk* in this case. This gives us good reason to think that the normic account of *risk* describes the core of the concept *risk*. That is, in typical cases, *risk* comes equipped with the intension and extension demarcated by the normic account. Crucially, this result is not dependent on the particular details of the marathon scenario, but rather on the need contained within it: for an agent to make a single decision so as to reduce disvalue under conditions of uncertainty. This is what makes the marathon case a typical case for *risk*.

As a final illustration, we can envision cases in which normic and probabilistic risk come apart. In these cases, we think that normic risk is the most useful concept for guiding an agent's decision-making, further suggesting that *normic risk* forms the explanatory core of the concept *risk*. For example, say that the agent knows that one of their parents has hypertrophic cardiomyopathy: then she has a 50% chance of having inherited this genetic condition (British Heart Foundation, 2023). Nevertheless, the overall probabilistic risk of suffering a cardiac arrest during the marathon would still be reasonably low. It is difficult to say with precision what this

suffering a fatal cardiac arrest is low *even if* one falls in the most at risk demographic (e.g. middle aged men who have run fewer than two marathons in the past, on Ghio et al's 2012 data), and so the probabilistic account may after all guide action effectively. This does not resolve the problem, though, that even calculating the most at risk statistic requires calculations that are impractical for many agents. Note further that we have here assumed that the agent has no symptoms, but the calculation becomes even more complex when we introduce a symptom such as fatigue or shortness of breath, both of which are symptoms of both hypertrophic cardiomyopathy and many other conditions (e.g. of COVID), several of which do not increase one's probabilistic risk of suffering a cardiac arrest during a marathon. Zooming out from the details of our central case, the general point is that it is difficult to see why a concept of *risk* which requires knowledge of, and calculation with, oftentimes complex and unobtainable probabilities would be the concept that has emerged for everyday deliberation aimed at reducing disvalue under conditions of uncertainty.

[8] That normalcy is relevant to individuals in a way that probabilities are not is a key part of Smith's resolution of a 'puzzle in evidence law' (2018, p. 1197), whereby purely statistical evidence that makes it more likely that a defendant is guilty than does non-statistical evidence nevertheless seems less probative of her guilt. Purely statistical evidence can never make it abnormal for a defendant to be innocent: if the only evidence we have against Joe is that he is part of a group of 100 people, 99 of whom committed a crime, 'he could simply be the one innocent person and *that's that*' (p. 1209). Smith thinks this explains why purely statistical evidence cannot suffice for conviction.

percentage is, but say that of the 50,000 people who run the London Marathon, 1000 of them have hypertrophic cardiomyopathy (in line with overall population averages of 2%). If 0.5 out of 50,000 people suffer a fatal cardiac arrest during or immediately after running a marathon (derived from the earlier probability of 1 in 100,000), then there is a 0.05% probability of one of the 1000 hypertrophic cardiomyopathy suffers having a fatal cardiac arrest during the marathon. This figure is undoubtedly higher than the 0.001% probability of fatal cardiac arrest that includes both people with underlying heart conditions and those without; but it is still a reasonably low probabilistic risk. The normic risk is higher: suffering from a fatal cardiac arrest caused by inherited hypertrophic cardiomyopathy during the marathon, knowing that one's parent has the condition, would not be particularly abnormal - it precisely would not cry out for an explanation.[9] The agent who knows that their parent has hypertrophic cardiomyopathy, then, can usefully guide her decision making by considering the marathon running to be reasonably high risk.[10]

3.3. The periphery of risk

Our central case is one in which one individual is trying to make a single decision so as to reduce disvalue for herself, in the short-term. In this case, *normic risk* is useful and exemplifies the concept of risk in its most explanatory basic form (viz., the core of *risk*). But this is not to say that there are not other, less core cases in which the very same function of *risk* would be better served by the notion output by a different theory. For example, there are cases in which agents need a concept to guide their decision-making with the aim of reducing disvalue not just for one individual over one decision, but for many individuals, over time and across situations. In order to be useful for these cases, the concept *risk* needs to become unmoored from the particular perspective of an agent making a decision aimed at reducing disvalue for herself, to become responsive to the needs of an agent making a decision where the aim is to reduce disvalue as it might occur across many individuals. In these cases, the concept of risk might pick out a different risk property. Call this axis 'de-individualisation', which measures the extent to which the agent's perspective in evaluating the risk of an event is tied to individual decisions. Highly individualised perspectives are concerned with evaluating risk in the course of guiding one-shot decisions that reduce disvalue for the agent; highly de-individualised perspectives are concerned with evaluating the risk of an event independently of how this might guide any particular

[9] This is not to say that *not* suffering from a fatal cardiac arrest in these circumstances would be abnormal: it would still be the case that not suffering a fatal cardiac arrest would not call out for an explanation.

[10] Note as a further illustration of how normic risk is responsive to bodies of evidence in a way that usefully guides agents' decision making, that it guides agents differently depending on their evidence. For on the probabilistic account, both the agents with 2% probability of having hypertrophic cardiomyopathy and the agent with 50% probability of having hypertrophic cardiomyopathy are guided to similar risk assessments regarding the risk of suffering from a fatal cardiac arrest during the marathon, provided that the overall probability of suffering from a cardiac arrest is sufficiently low. In contrast, the normic account is responsive to whether the agents' evidence make it such that an explanation would be called for if the risk event would obtain, leading to different - and more useful - risk assessments.

decisions (see §4); and in between, there are perspectives concerned with evaluating risk in the course of guiding decision-making to reduce disvalue across multiple agents or situations.

We can see this process as, in one sense, mirroring the logic of the process of objectivisation (Craig 1999). During objectivisation, a concept moves from exclusively serving a function that is tied to the particular needs of one individual - what Craig calls the "subjectivist stance" (p. 83) - to serving a function that is tied to the broader needs of an entire community. Craig argues that the ancestor of the concept of *knowledge* serves the function of flagging "an informant who is satisfactory for my purposes, here and now, with my present beliefs and capacities for receiving information" (p. 85). The process of objectivisation pushes this original concept into serving the function of flagging an informant who is reliable enough "whatever the particular circumstances of the inquirer, whatever rewards and penalties hang over him and whatever his attitude to them" (p. 91). In a (partially) structurally analogous way, we posit that the concept of *risk* moves from serving the function of guiding decision making in one-shot decisions that aim to reduce disvalue in a situation with a single individual to doing so across situations with multiple individuals. However, unlike the process of objectivisation, in de-individualisation, the core individual case that explains why the concept emerges need not fade from use. In our central case, quite the opposite occurs: the individual case remains the explanatory core of the concept *risk*. What the de-individualisation axis tracks is modifications to the concept in specialised and particular cases that explain its fulfilling its function in these specialised and particular cases. The structural analogy to objectivisation, then, is partial: whereas objectivisation results in the implementation of the revised concept across the board, de-individualisation results in the implementation of the revised concept in specific circumstances where particular needs are present, leaving the original concept operative for everyday circumstances that feature the everyday needs that explain the core use of the concept.

To illustrate, imagine a member of the City of Edinburgh Council taking part in a deliberation over whether Edinburgh should hold a marathon. What matters to her *qua* decision-maker is not the risk of any particular person dying as a result of running the marathon. Rather, her concerns are at the level of the marathon-running population in general. The most pressing risk-events are the deaths of *any* runners. She is concerned with whether the risk of such events obtaining is too high to justify holding the marathon, whether and how this risk can be lowered, and so on. What must *risk* look like to be helpful for this decision-maker?

Normic risk issues the verdict that there is a low-risk of any given runner dying from cardiac arrest. People tend not to run marathons if they know they have heart conditions that would lead them to suffer cardiac arrest under sufficient exertion. Then for each runner, relative to her evidence, some explanation would be required if she were to die as a result of running the marathon. But if the normic risk of any particular runner dying from cardiac arrest is low, then the normic risk of some runner dying from cardiac arrest is likewise low. This is because the normic risk of a disjunction is only as high as the normic risk of its most normal disjunct. The most normal world in which $A \vee B$ is true is a world in which A is true or a world in which B is true. Then the most normal world in which some runner dies from cardiac arrest is a world in which either Runner 1 dies from cardiac arrest, or in which Runner 2 dies from cardiac arrest, or in which Runner 3 dies from cardiac arrest, and so on. But each of these worlds is abnormal.

Then the most normal world in which some runner dies from cardiac arrest is abnormal, and so the normic risk of some runner dying from cardiac arrest is low.

But despite the low normic risk, it is nevertheless the case that runners do die from cardiac arrest either during or immediately after running marathons: approximately one in 100,000. Assuming that the Edinburgh Marathon would have around 10,000 participants, there is a 0.1 probability that some runner will die from cardiac arrest. This is what should concern the council member. That any runner's death would be abnormal is irrelevant for her decision-making, as abnormal events can and do occur. The council member cares about reducing the frequency with which events of this kind occur, whether or not these occurrences are normal. What the council member cares about, then, is *probabilistic risk*.

This point is further highlighted by considering the steps that the council member might take to try and mitigate the risk of any marathon runner dying from a fatal cardiac arrest. For example, Kim at al's 2012 review found that a high percentage (88%) of marathon runners who survived cardiac arrests received automated defibrillator assistance at the scene (compared with non-survivors, of which only 35% received automated defibrillator assistance at the scene). Then, the council member can use the fact that her evidence indicates that the presence of automated defibrillators reduces the risk of marathon runners dying of cardiac arrest to guide her decision making: for example, she might invest council funds in purchasing automated defibrillators, and train volunteers in their use. In contrast, it is unclear whether the normic account of risk generates this clear action-guiding result. For although there is a sense in which death by cardiac arrest is more normal in the absence of automated defibrillator assistance than in the presence of automated defibrillator assistance, no special explanation is called for in cases where automated defibrillator assistance does not result in resuscitation. More generally, when it comes to reducing disvalue at the level of populations, whether a disvaluable event occurring would call out for explanation is less important than how frequently events of that kind occur. Therefore, when making decisions with the aim of reducing disvalue not just for one individual, here and now, but for some broader population, over time and across situations, we should appeal to *probabilistic risk*. In its periphery, *risk* is demarcated as the probabilistic account has it.

4. The outer periphery of risk

We can follow this line of de-individualisation to imagine what a fully de-individualised concept *risk* would look like. The normic and probabilistic accounts both relativise *risk* to a body of evidence. A fully de-individualised concept *risk* would not be relativised to any body of evidence; or rather, it would be relativised to the maximally inclusive body of evidence that consists of all and only the facts about the world. This is not a body of evidence possessed by any actual agent or group of agents.

This fully de-individualised concept *risk* would capture the sense in which an event can have some level of risk given the totality of facts, irrespective of any particular agent's perspective. This is what Pritchard's modal account of *risk* aims to capture. On the modal account, *risk* is determined by how close is the closest world in which a risk-event obtains. Whether a world is

close is not a matter of whether anyone's evidence suggests that it is close. Rather, it is determined by what the actual world is like, regardless of whether anyone knows that the actual world is that way. Then *modal risk* is independent of any agent's evidence. The modal risk of some marathon runner suffering from a fatal cardiac arrest is determined solely by what the actual world is like, and in particular whether the actual runner has an underlying heart condition such as hypertrophic cardiomyopathy, not by whether the evidence of any would-be marathon runner affords reason to think that she has such a condition, nor by facts about what percentage of runners suffer fatal cardiac arrests during or immediately after a marathon.

Given our hypothesis about the function of *risk*, this fully de-individualised form of *risk* might seem entirely non-functional. After all, decision-makers are limited by their bodies of evidence, and no fallible human decision-maker has as her body of evidence the totality of facts. But we can make sense of the idea of a risk that is independent of any body of evidence, and imagine scenarios in which it is useful to think in these terms. For example, if there is an asteroid heading towards Earth, there is a sense in which the asteroid crashing into Earth is high-risk, even if nobody has any evidence suggesting that the asteroid exists, nor that Earth has been hit by similar asteroids in the past. But if no one has this evidence, then it is not probabilistically or normically high-risk, relative to anyone's evidence, that this asteroid is going to crash into Earth. The idea that some event that nobody knows anything about can nevertheless have some positive level of risk can only be made sense of on a concept *risk* that is not relativised to bodies of evidence. Insofar as it is useful to be able to think in these terms, *modal risk* is useful, even for fallible and evidence-constrained creatures like us.

Further, in the specialised context of philosophical inquiry, this de-individualised *risk* concept is often applicable. Philosophers are concerned with, among other things, discovering objective truths about the world. *Modal risk* picks out an objective property of the world: how much would have to be different to get to a world in which the relevant disvaluable event occurs. It is easy to see why philosophers would have use for this fully de-individualised form of the concept. Thus we suggest that, in its fully de-individualised periphery, *risk* is demarcated by the intension and extension that the modal account of risk posits.

There are, therefore, situations that call out for the use of *modal risk* rather than *normic* or *probabilistic risk*. However, it is a problem for Pritchard that *modal risk* is functional only in peripheral cases. Pritchard motivates his shift from anti-luck to anti-risk epistemology by appealing to the "strategic" value of *risk* over *luck*. Pritchard argues that luck is "essentially backwards-looking", in that we make judgements about luck only after an event has obtained, while *risk* has a "forward-looking dimension" such that we make risk judgements about events that have not yet occurred" (2022, p. 16; see also Navarro 2019, p. 69). Consider an example. If you walk across a rickety bridge over a ravine without falling in, you were *lucky* to get to the other side unscathed; but if you are about to cross said bridge, you would judge that you would be at *high risk* of falling into the ravine, were you to do so. This makes *risk* more useful than *luck* from the perspective of someone trying to decide what to do (rather than evaluating whether she ought to have done something she already did): whether she should, for example, trade-off risks of one kind for risks of another, or whether avoiding risk is worth the costs of doing so (Pritchard 2022, p. 20). But we have shown that *modal risk* simply does not have this strategic value. An

agent typically has reason to think in terms of *risk* precisely because she does not know many details of her situation. In these cases, she lacks epistemic access to which worlds are close, thus cannot appeal to *modal risk* in the ways Pritchard recommends: for example, she cannot think about trading off risks of one kind for risks of another, as she cannot determine any of these risks. Thus *modal risk* lacks the strategic value that Pritchard appeals to in motivating his move from anti-luck to anti-risk epistemology.

To conclude our exposition of the core-to-periphery structure that we have posited for *risk*, we wish to emphasise two points. First, the core-to-periphery structure reflects an explanatory structure: it explains why the concept has the shape that it has in different circumstances. *Normic risk* is core in the sense of being the shape that the concept *risk* takes in everyday cases (viz., in cases that explain why we have the concept to begin with); *probabilistic risk* is at the periphery in the sense of being the shape that the concept *risk* takes in the specialised case of an agent making a decision so as to reduce disvalue across many individuals (viz., in cases that explain why we use the concept of *risk* in these multi-individual cases); and *modal risk* is at the outer periphery in the sense of being the shape that the concept risk takes in very particular philosophical contexts, where what matters is an objective property of the world (viz., in cases that explain why we use the concept of *risk* in these philosophical contexts). But to say that a particular *risk* concept is at the periphery or outer periphery is not to suggest that it is somehow deficient as a concept of *risk* - it is rather that the cases which explain why we have *that* concept of risk sit at the periphery of the cases that explain why we use the concept of *risk* in general. And *vice versa*, to say that a particular *risk* concept is core is just to say that it sits at the core of the cases that explain why we have the concept of *risk* in general.

The second point to emphasise is that the individualisation axis that we have posited as underpinning this structure is not incidentally instantiated in each of the three risk accounts, but takes a particular form in each one. That is to say, it is not merely that *risk* is most functional in everyday cases when it is relativised to evidence, *any which way*; but more substantially, that *normic risk* is relativised to evidence in ways that make it useful for creatures like us, given our abilities, prior knowledge and concerns. Likewise, it is not just that in the periphery, probabilistic risk is functional purely on account of being semi-individualised - it is rather than probabilistic risk is semi-individualised in the right way to be useful given the abilities, prior knowledge and concerns of an agent making a decision from the perspective of aiming to reduce disvalue across multiple individuals. Thus, the reverse-engineering project that we've undertaken shows not just that *risk* is de-individualised at its core, semi-individualised in the periphery, and completely de-individualised in the outer periphery; but more substantially that risk is *normic risk* at its core, *probabilistic risk* in the periphery, and *modal risk* in the outer periphery.

5. Principled Pluralism

Our conceptual reverse-engineering analysis reveals that *risk* is a pluralist, rather than monist, concept. *Risk* is "both one and many" (Lynch 2009, p. 69): unified in serving one function, but taking distinct forms in different situations to meet this same function. *Risk*'s one function is best served by the notion of *risk* issued by the normic account in typical cases involving individual decision-makers; by that issued by the probabilistic account where our concern is reducing the frequency of some kind of risk-events; and by the fully de-individualised notion of *risk* issued by the modal account in philosophical contexts.

Ebert, Smith and Durbach (2020) also argue for *risk* pluralism, writing that "more research will be required" (448) to develop their proposal.[11] Our analysis takes up this call, and has the additional advantage of being a principled pluralism. First, Ebert et al.'s pluralism is motivated by pessimism regarding the stagnating *risk* literature, but their methodology is of a piece with that debate: they simply note that pluralism predicts the competing intuitions generated by the alleged counterexamples. In contrast, we take the stagnation as a cue to change methodology, and this new methodology generates a pluralist picture. Second, we explain why *risk* takes a pluralist form to begin with: to best serve the function of guiding decision making so as to reduce disvalue under conditions of uncertainty. Third, we bring order to *risk*'s multiple forms, outlining how they relate to one another and hang together: *risk* has a core-to-periphery structure, taking a different form in typical ('core') cases than it does in less typical ('peripheral') cases.

Our principled pluralism confers two key advantages over monists accounts, which we now discuss: it explains the proliferation of counter-examples that led to the impasse in the risk debate, and it makes sense of the otherwise puzzling distinction between subjective and objective risk.

5.1. Counter-Examples

Our reverse-engineered pluralist concept *risk* affords a principled explanation for the counterexamples which lead to the epistemic risk debate impasse. In a nutshell, the cases cease to be counterexamples once we see which notion of risk is at play in each case. For example, our pluralist account predicts that Pritchard's analysis of the bomb case is correct: in this case, the relevant notion of *risk* is that demarcated by the modal account, according to which Bomb 1 is at higher risk of detonating than Bomb 2, despite their identical probability of detonating. This is because the case is presented to test the intuitions of an uninvolved audience who cannot

[11] Bricker (2018) also argues that the variations in empirical observations on risk judgments, including empirical data suggesting that risk judgements are produced by multiple cognitive systems, "seems to defy any monistic description" (p. 207).

intervene in the events that trigger the bomb but who are asked, from a third party perspective, to evaluate the levels of risk presented in each case. Therefore, the risk invoked in the Bomb case sits at the outer periphery, where risk is non-action-guiding. In contrast, the risk property in the West Australian House case is a paradigmatic instance of the core of *risk*: an agent is thinking about the risk of their walls containing asbestos in order to decide whether to drill into the walls. In this case, the agent needs *risk* to be action-guiding, so the notion of *risk* given by the normic account is most fitting.

Our pluralist picture of risk, then, has the added advantage of predicting which account of *risk* best captures practices in each individual context, by reflecting on what specialised needs are present in that context. This goes some way towards answering what Ebert, Smith and Durbach call the "meta-normative issue" of determining which form of *risk* ought to be used in a given context (2020, p. 449). The form that *risk* should take is that which is best suited to guide decision-making under uncertainty to reduce disvalue, given the particular needs of the context.

5.2. Subjective and Objective risk

The pluralistic account of *risk* that we develop in §3 can be used to resolve a long-standing problem in risk-analysis: how to distinguish between subjective and objective risk. The subjective/objective risk distinction is standardly thought to be crucially important for the study of *risk* (see for example Bradbury 1989, p. 389; Möller 2012), yet is poorly characterised. Sven Ove Hansson suggests that an objective account of *risk* includes "(only) objective facts about the physical world" (2010, p. 232), while a subjective account of *risk* "does not refer to any objective facts about the physical world" (233). Hansson's way of marking the distinction is inspired by Harry Otway and Kerry Thomas, who themselves endorse a subjective account, according to which *risk* is "a subjective experience (or a future projection of an experience) which is meaningful for, and can be thought about, judged and felt by anyone, expert or layperson", not an objective fact about a world that exists independently of our subjective experience (1982, p. 69-70).

As Hansson himself argues, both objective and subjective accounts of *risk*, thus characterised, are obviously false. As *risk* involves disvalue, and values often cannot be characterised fully in terms of objective facts about the physical world, independent of our experience, then any account of *risk* that makes appeal only to objective facts about the physical world (i.e. any objective account of *risk*, on Hansson's definition) is obviously false. But as *risk* has a factual component – for example, if you risk losing your leg if you tread on a landmine, it must be the case that landmines tend to dismember people who tread on them – any account of *risk* that makes no appeal to facts about the physical world (i.e. any subjective account of *risk*, on Hansson's definition) is even more obviously false. Hansson takes this as reason to endorse his

"dual risk thesis", according to which risk is neither fully objective nor fully subjective; rather, any "accurate and reasonably complete characterization of a risk must refer both to objective facts about the physical world and to (value) statements that do not refer to objective facts about the physical world" (236). But why not instead think that this shows that Hansson's characterisation of the subjective/objective risk distinction is faulty? After all, doesn't charity demand that we put the distinction in a way that does not make both theses obviously false?

We suggest that we can think of *subjective risk* as the form the concept *risk* takes in its explanatory core. That is, *subjective risk* is the concept at work when a subject is making a decision under conditions of uncertainty with the aim of reducing disvalue. *Objective risk* is the form the concept *risk* takes after it has been de-individualised in response to further needs. *Subjective risk* could take the form proposed by either the probabilistic or the normic accounts of *risk*, which both have it that risk is determined relative to a subject's or group's evidence. *Objective risk* can take the form proposed by the normic and probabilistic accounts, relative to a maximal body of evidence; or that proposed by the modal account, where *risk* involves no evidence-relativity at all.

6. Applied to epistemic risk

We will now apply the foregoing discussion to *epistemic risk*. Recall that the phrase 'epistemic risk', as found in the epistemological literature, has two senses. In the first sense, *epistemic risk* is simply a kind of *risk*: what makes it epistemic is that it concerns epistemic disvalue. In the second, more restrictive sense, 'epistemic risk' picks out the risk of false belief. This more restrictive sense of 'epistemic risk' is found more often in the literature (see Collins 1996, Wright 2004, Lasonen-Aarnio 2008, Smith 2012, Pritchard 2016), though it is becoming increasingly popular to see more expansive uses of 'epistemic risk' (see for example Navarro 2021 and Pritchard 2021).

Everything we've said about *risk* applies to *epistemic risk*, as picked out by the first sense of 'epistemic risk'. If *risk* functions to guide decision-making so as to reduce disvalue under conditions of uncertainty, then *epistemic risk* functions to guide decision-making so as to reduce epistemic disvalue under conditions of uncertainty. Epistemically disvaluable events include losing epistemic goods, such as true belief, knowledge or understanding; failing to obtain epistemic goods; or acquiring epistemic bads, such as false belief, misleading evidence or misunderstanding.

As *epistemic risk* in this sense is just a kind of *risk*, we see the same core-to-periphery structure in *epistemic risk* that we saw in *risk*. The core case is one in which an agent is trying to make a decision that reduces epistemic disvalue under conditions of uncertainty. Consider such an agent,

who is deliberating about whether P. It would be disvaluable for her to form a false belief whether P, and it would be disvaluable for her to fail to form a true belief whether P. How disvaluable each outcome is will vary. For example, if she needs to form a belief about whether *this is the right train to catch*, forming a false belief and failing to form a true belief are equally disvaluable: both lead her to miss her train. In any case, the *epistemic risk* concept useful for this inquirer is that issued by the normic account. She should ask: given my evidence, could it just so happen that in forming a belief that (say) P, I would form a false belief? If this could just so happen – if her forming a false belief that P, given her evidence, would not call out for explanation – then she ought not form a belief that P. She should also ask: given my evidence, could it just so happen that in failing to form a belief that P, I would miss out on true belief? If this could just so happen, then she ought to form a belief. Suppose that the inquirer is deliberating about whether it is raining. She looks out the window and sees what looks like rain falling. Given her evidence, a belief *that it is raining* being false would call out for explanation: if it isn't raining, why does it look like it is? Further, if she were to fail to form a belief *that it's raining* on the basis of this evidence, more explanation would be required if she thereby failed to form a false belief than if she thereby failed to form a true belief: again, if it isn't raining, why does it look like it is? Then forming a belief that it's raining is her best option for reducing epistemic disvalue: for avoiding both false belief and the missed opportunity for true belief.

The probabilistic notion of *risk* is less useful for guiding action in this situation. The inquirer is not interested in how many relevantly similar beliefs to that which she would form, based on relevantly similar evidence, would be true and how many false; rather, she is interested in whether this particular belief would be true or false. She is not, then, interested in the probability of forming a false belief or the probability of failing to form a true belief. However, the probabilistic notion of *epistemic risk* usefully guides decision-making when an agent is concerned with reducing the frequency of some kind of epistemic disvalue. Consider a government education minister who is deciding which educational policies to implement, with an eye to reducing lack of understanding in the nation's pupils. Whether a given policy would make it more or less abnormal for an individual pupil's understanding to increase will depend on the pupil's circumstances, learning style and interests. However, the education minister is not interested in reducing some individual pupil's misunderstanding, but in reducing the frequency of misunderstanding across many pupils. For this purpose, she can fruitfully look to data on whether relevantly similar policies have correlated well with reduced misunderstanding over a population.

In both of these cases, *epistemic risk* is relativised to a body of evidence. But just as for *risk* in general, we can imagine cases that call out for the use of a fully de-individualised *epistemic risk* concept. For example, we can ask of some epistemic good, irrespective of any body of evidence, how easily some subject might miss out on it: how close is the closest world in which, say, the

detective looks over some crucial piece of evidence? When we use the concept *epistemic risk* to ask these kinds of questions, it picks out the property of modal epistemic risk: epistemic disvalue in close worlds. Fully de-individualised *epistemic risk*, then, is the modal account's notion of *epistemic risk*.

The modal account's notion of *epistemic risk* is not helpfully action-guiding in typical cases in which agents appeal to *epistemic risk*, as these are cases in which the agent is ignorant of many details of her situation, so doesn't know what the actual world is like and by extension doesn't know which worlds are close. But this notion of *epistemic risk* is useful for epistemologists. Epistemologists will generally evaluate subjects' epistemic positions from what Bernard Williams (1973, p. 146) calls the "examiner situation": the situation in which the epistemologist knows that P, knows that S believes that P, knows all the relevant facts about S's situation, and is determining whether S knows that P. The epistemologist can determine how close is the closest world in which, for example, a subject's true belief is false, because she knows all the relevant facts about the subject's situation that determine world ordering. If the world in which a subject forms a false belief is close, then the epistemologist says that forming this belief is high-risk, irrespective of the subject's evidence – in particular, irrespective of whether the subject's evidence makes it the case that false belief is unlikely, or abnormal. Thus the modal account's notion of *epistemic risk* is useful to the epistemologist. In particular, this is the notion of *epistemic risk* that must underpin anti-risk epistemology, according to which knowledge is incompatible with high levels of epistemic risk (Pritchard 2015, 2016; Navarro 2019, 2021), since anti-risk epistemology is concerned with what it takes for S's belief that P to be knowledge, not with whether some (limited) body of evidence suggests that S's belief that P constitutes knowledge.

That epistemologists appeal to *epistemic risk* primarily in the context of assessing whether a true belief constitutes knowledge explains why we find the second sense of 'epistemic risk', on which it refers only to the risk of forming a false belief, more often in the literature than we find the first. When an inquirer is using *epistemic risk* to guide her inquiry, she must consider different kinds of epistemic risks, and weigh up the importance of reducing risks of each kind. For example, she must consider whether forming a false belief would be worse than missing out on true belief; or whether forming a belief is sufficiently important that failing to form any belief is as bad as forming a false belief. But from the examiner situation, all that one is concerned with is whether a subject's true belief suffices for knowledge; this is determined by how close are the closest worlds in which the subject's belief is false. Thus there is only one kind of epistemic risk that is relevant: the risk of forming a false belief.

We take it to be an advantage of our pluralist account of *risk* that *epistemic risk* can be understood as a subset of *risk*. *Epistemic risk* is of the same kind as *risk simpliciter*; what is

distinctive about *epistemic risk* is that its risk-events are epistemic risk-events. This makes ours a simple, parsimonious theory. But we can nevertheless explain why 'epistemic risk' is used, on the face of it somewhat esoterically, to pick out a very narrow kind of epistemic risk-event: the event of a subject's forming a false belief. As such, we can make sense of the phrase 'epistemic risk' as most commonly found in the epistemological literature. So the simplicity of our theory does not compromise its general applicability.

7. Conclusion

We have reverse-engineered *risk*, generating an account of *risk* as a pluralistic concept with a core-to-periphery structure. On our account, *risk* takes the form demarcated by the normic account at its explanatory core, that demarcated by the probabilistic account after a partial process of de-individualisation, and that demarcated by the modal account after full de-individualisation. As such, our conceptual reverse-engineering project vindicates all three accounts of *risk*: the probabilistic account, modal account and normic account each articulate a form that *risk* takes at some point in the core-to-periphery structure. Our pluralist account of *risk* improves on Ebert, Smith and Durbach's *risk*-pluralism by explaining the 'why and how' of *risk*-pluralism: why we would expect *risk* to vary in content in different contexts, and how it can remain the same concept, given this variety. We expect *risk* to vary in content in different contexts as this is what is required to serve its function in those contexts. *Risk* can so vary in content because of its core-to-periphery structure: it is normic at its core, probabilistic at its inner periphery, and modal at the outer periphery. The periphery is understood as an elaboration of the core, serving the same function, and as such, as part of one and the same concept. We drew attention to further advantages of our account of *risk*, such as its ability to distinguish objective from subjective risk. Finally, we applied this picture of *risk* to the epistemic realm, to explain an ambiguity in the epistemological literature on *epistemic risk*.

References

Backes, Marvin 2018: 'Normalcy, justification, and the easy-defeat problem'. Philosophical Studies, 176(11), pp. 2877-95.

Blome-Tillmann, Michael 2009: 'Knowledge and Presuppositions'. Mind, 118(470), pp. 241-94.

Bradbury, Judith 1989. 'The Policy Implications of Differing Concepts of Risk'. Science, Technology, and Human Values, 14, pp. 380-399.

British Heart Foundation 2023: Hypertrophic cardiomyopathy, online at https://www.bhf.org.uk/informationsupport/conditions/cardiomyopathy/hypertrophic-cardiomyopathy (Accessed 15th September 2023).

Bricker, Adam 2018: ‘Do Judgements about Risk Track Modal Ordering?’. Thought, 7, pp. 200-8.

Buchak, Lara 2017: ‘Taking Risks Behind the Veil of Ignorance’. Ethics, 127(3), pp. 610-644.

Collins, Arthur W. 1996: ‘Moore’s Paradox and Epistemic Risk’. The Philosophical Quarterly, 46(184), 308-19.

Craig, Edward 1990: Knowledge and the state of nature: An essay in conceptual synthesis. Oxford: Oxford University Press.

Ebert, Philip, Smith, Martin. and Durbach, Ian 2020: ‘Varieties of Risk’. Philosophy and Phenomenological Research, 101(2), pp. 432-55.

Elsom, Derek 2001: ‘Deaths and Injuries Caused by Lightning in the United Kingdom: Analyses of Two Databases’. Atmospheric Research, 56(1-4), pp. 325-34.

Fratantonio, Giada 2021: ‘Evidence, Risk, and Proof Paradoxes: Pessimism about the Epistemic Project’. The International Journal of Evidence & Proof, 25(4), pp. 307–325.

Fricker, Miranda 2008: ‘Scepticism and the Genealogy of Knowledge: Situating Epistemology in Time’. Philosophical Papers, 37(1), pp. 27-50.

——— 2016: ‘What’s the Point of Blame? A Paradigm-Based Explanation’. Noûs, 50(1), pp. 165-83.

Gardiner, Georgi 2015: ‘Teleologies and the Methodology of Epistemology’. In J. Greco and D. Henderson (eds.), Epistemic Evaluation: Purposeful Epistemology, pp. 31-44. Oxford: Oxford University Press.

Ghio, F. E., Pieri, M., Agracheva, A., Melisurgo, G., Ponti, A., & Serini, C. 2012: ‘Sudden cardiac arrest in a marathon runner. A case report.’ *HSR proceedings in intensive care & cardiovascular anesthesia*, 4(2), 130–132.

Grimm, Stephen R. 2015: ‘Knowledge, Practical Interests, and Rising Tides’. In J. Greco and D. Henderson (eds.), Epistemic Evaluation: Purposeful Epistemology, pp. 117-35. Oxford: Oxford University Press.

Hannon, Michael 2019: What’s the Point of Knowledge? A Function-First Epistemology. Oxford: Oxford University Press.

Hansson, Sven Ove 2010. ‘Risk: Objective or Subjective, Facts or Values’. Journal of Risk Research, 13(2), 231-238.

Hume, David 1739 [2014]: ‘Of the origin of justice and property’. *A Treatise of Human Nature*, in L. A. Selby-Bigge, and P. H. Nidditch (eds). Second edition. Oxford: Oxford University Press.

Kaplan, Stan 1997: ‘The Words of Risk Analysis’. *Risk Analysis*, 17 (4), pp. 407-417.

Kim, Jonathan. H., Malhotra, R., Chiampas, G., d'Hemecourt, P., Troyanos, C., Cianca, J., Smith, R. N., Wang, T. J., Roberts, W. O., Thompson, P. D., Baggish, A. L., & Race Associated Cardiac Arrest Event Registry (RACER) Study Group 2012: ‘Cardiac arrest during long-distance running races’. *The New England journal of medicine*, 366(2), 130–140.

Lasonen-Aarnio, Maria 2008: ‘Single Premise Deduction and Risk’. Philosophical Studies, 141, pp. 157-173.

Lee-Stronach, Chad 2018: ‘Moral Priorities Under Risk’. Canadian Journal of Philosophy, 48(6), pp. 793-911.

Lewis, David 1973: Counterfactuals. Oxford: Blackwell.

Lynch, Michael 2009: Truth as One and Many. Oxford: Oxford University Press.

Möller, Niklas 2012. ‘The Concepts of Risk and Safety’. In S. Roeser, R. Hillerbrand, P. Sandin and M. Peterson (Eds.), Handbook of Risk Theory: Epistemology, Decision Theory, Ethics, and Social Implications of Risk.

Navarro, Jesús 2019: ‘Luck and Risk’. Metaphilosophy, 50(1-2), pp. 63-75.

——— 2021: ‘Epistemic Luck and Epistemic Risk’. Erkenntnis.

Newton, Lilith 2022: ‘Epistemic anxiety and epistemic risk’. *Synthese,* 200 (4), pp. 1-23.

Otway, Harry and Thomas, Kelly: 1982. ‘Reflections on risk perception and policy’. Risk Analysis, 2(2), 69-82.

Pritchard, Duncan 2015: ‘Risk’. Metaphilosophy, 46, pp. 436-461.

——— 2016: ‘Epistemic Risk’. Journal of Philosophy, 113, pp. 550-571.

——— 2022a: ‘In Defence of the Modal Account of Legal Risk’. Synthese 200(290). DOI: 10.1007/s11229-022-03693-z

——— 2022b: ‘Varieties of Epistemic Risk’. Acta Analytica, 37, pp. 9-23.

Queloz, Matthieu 2021: The Practical Origin of Ideas. Oxford: Oxford University Press.

Smith, Martin 2012: ‘Two Notions of Epistemic Risk’. Erkenntnis, 78, pp. 1069-1079.

——— 2016: Between Probability and Certainty: What Justifies Belief. Oxford: Oxford University Press.

——— 2018: ‘When Does Evidence Suffice for Conviction?’ Mind, 127(508), pp. 1193-1218.

Thoma, Johanna 2019: ‘Risk Aversion and the Long Run’. Ethics, 129(2), pp. 230-253.

Thomasson, Amie 2020: ‘A Pragmatic Method for Normative Conceptual Work’. In A. Burgess, H. Cappelen and D. Plunkett (eds.), Conceptual Engineering and Conceptual Ethics. Oxford: Oxford University Press.

Thomson, Judith Jarvis 1983: ‘Some Questions about Government Regulation of Behavior’. In T. Machan and M. Johnson (eds.), Rights and Regulation: Ethical, Political and Economic Issues. Cambridge, MA: Ballinger.

Waite, Oliver, Andy Smith, Luke Madge, Hannah Spring and Nathalie Noret 2016: ‘Sudden Cardiac Arrest Death in Marathons: A Systematic Review’. The Physician and Sportsmedicine, 44(1), pp. 79-84.

Williams, Bernard 1973: ‘Deciding to Believe’. In his Problems of the Self: Philosophical Papers 1956-1972. Cambridge: Cambridge University Press.

——— 2002: Truth and Truthfulness: An Essay in Genealogy. Princeton, NJ: Princeton University Press.

Wright, Crispin 2004: ‘Warrant for Nothing (and Foundations for Free?)’. Aristotelian Society Supplementary Volume, 78(1), pp. 167-212.