# Catastrophic Risk

H. Orri Stefánsson[*] [†]

Final version forthcoming in *Philosophy Compass*.

**Abstract**

Catastrophic risk raises questions that are not only of practical importance, but also of great philosophical interest, such as how to define catastrophe and what distinguishes catastrophic outcomes from non-catastrophic ones. Catastrophic risk also raises questions about how to rationally respond to such risks. How to rationally respond arguably partly depends on the severity of the uncertainty, for instance, whether quantitative probabilistic information is available, or whether only comparative likelihood information is available, or neither type of information. Finally, catastrophic risk raises important ethical questions about what to do when catastrophe avoidance conflicts with equity promotion.

Keywords: *Catastrophe; Risk; Rationality; Equity; Tradeoffs.*

[*]Stockholm University, Swedish Collegium for Advanced Study, and Institute for Futures Studies. Email: orri.stefansson@philosophy.su.se; website: www.orristefansson.is.

[†]I have benefited from very useful comments from Richard Bradley, Martin Peterson, an anonymous referee, and from subject editor Guy Fletcher.

# 1 Introduction

The 2010 Kaprun train accident killed 155 people, making it Austria's deadliest ever traffic accident. The country observed two days of national mourning, and the regional governor referred to the accident as a "catastrophe".[1] That same year, there were 552 road traffic fatalities in Austria.[2] Few would call that a catastrophe; on the contrary, the steady decrease in annual road traffic fatalities had been seen as cause for celebration.

Why is it that 155 people dying in a train accident is a catastrophe, while more than three times as many people dying in road traffic accidents in one year is not a catastrophe? Is it worse, from a moral or social point of view, that 155 people die in one accident, rather than, say, 155 people dying in one hundred accidents spread over four months? And should society respond differently to a risk that is expected to kill some number of people in multiple incidents spread over the next four months, say, than to a risk that is expected to kill the same number of people in one incident some time during the next four months?

This article will discuss the above and some further philosophical questions concerning the treatment of catastrophic risk. For instance, I shall consider whether catastrophic risk either permits or requires us to violate principles of rationality that we are required to satisfy in more ordinary choice situations. Moreover, I shall discuss tradeoffs between increased equity and reduced catastrophic risk. But first, I discuss different ways in which

[1] The governor was cited in a BBC story about the accident: http://news.bbc.co.uk/2/hi/europe/3502265.stm.

[2] The figures are taken from https://www.statista.com/statistics/437842/number-of-road-deaths-in-austria/.

"catastrophe" and "catastrophic risk" have been defined.

# 2 Defining "catastrophe"

## 2.1 Risk and catastrophic risk

A *risk*, as I shall be using the term, is a (greater than zero) probability[3] of an undesirable outcome, that is, an outcome that is by the relevant decision-maker considered to be worse than the status quo. To put it slightly more formally: I shall be thinking of a risk as a pair, consisting of a positive probability and an undesirable outcome.

Both the probability and the potential undesirable outcome may be either known or unknown. For instance, playing Russian roulette with one out of six chambers loaded imposes a risk where both the probability (1/6) and the potential undesirable outcome (death) are known. Smoking, on the other hand, comes with a risk where the potential undesirable outcome is known (lung cancer) but the exact probability for each individual is not known (even though one can make pretty good predictions, even at the individual level, based on statistical data). Finally, geo-engineering, and radically new technologies, may bring with them risks where both the probabilities and the potential undesirable outcomes are unknown.

A *catastrophic* risk, as I shall use the term, is simply a (greater than zero) probability of a catastrophic outcome. Sometimes an outcome is only catastrophic for some individual. For instance, a toddler's death would (typically)

[3] A "probability" is a real number between zero and one, that satisfies the following additivity property: if $A$ and $B$ are mutually incompatible, then the probability of $A$ or $B$ obtaining is the probability of $A$ obtaining plus the probability of $B$ obtaining.

be a catastrophe for its parent. From a social point of view, however, a single child's death would (typically) not be a catastrophe. Other outcomes are catastrophic for a country, or some region. For instance, the Kaprun disaster was, we can assume, a catastrophe for Austria, or at least for the state of Salzburg. But 155 deaths in a train accident is not a global catastrophe. The Spanish flu, on the other hand, which is thought to have killed 50 to 100 million people around the world, could reasonably be called a global catastrophe.

## 2.2 Catastrophe and utility threshold

Do different catastrophes, and different types of catastrophes, have anything in common that we could use to define the term, "catastrophe", more precisely? One very natural idea, employed by Peterson (2002a), is to first construct a measure of the desirability of various outcomes, according to the agent of interest,[4] and then define a catastrophe as any outcome that is assigned a lower value than some threshold, by this measure of desirability.

To make this idea more precise, suppose that $u_i$ is a *utility* measure, that represents how (un)desirable agent $i$ considers different outcomes. For instance, $u_i(o) = m$ means that agent $i$ considers outcome $o$ (un)desirable to degree $m$. We can formulate the the simplest definition of "catastrophe" that Peterson considers as follows:[5]

**Definition** (Peterson on catastrophe)**.** *For each utility function, $u_i$, there is some*

[4]Note that the "agent" could be a group, or a society, which is then assumed to have a collective preference that can be represented by some measure of desirability (i.e., a utility function).

[5]The slightly more complex definitions that he considers concern distances from either the highest utility that is possible in a choice situation or the highest expected utility on offer.

*value $n$, such that an outcome $o$ is catastrophic, according to $i$, just in case $u_i(c) \leq n$.*[6]

This definition shares some similarities with a definition used by some economists, where a catastrophe is defined as an event that takes consumption below some threshold, or reduces consumption by some particular magnitude (see e.g. Martin and Pindyck 2015).

How does this way of defining "catastrophe" handle the opening example of this article? Consider for instance the comparison between on the one hand 155 deaths in a single train crash and, on the other hand, 155 deaths (or more) from one hundred road traffic accidents spread over four months. Assuming that the train crash is a (local) catastrophe, we could set the threshold utility, $n$, such that the train crash is assigned a utility lower than $n$ (according to, say, the utility function of the Austrian "social planner"). In contrast, I surmise that each individual road traffic accident is not, from the societal perspective, considered a catastrophe. But what does the definition imply for the collection of 155 deaths from multiple traffic accidents? That would depend on whether we count these deaths as a single outcome (since "catastrophic" defined relative to *an* outcome).[7]

Suppose, for instance, that a decision to increase the speed limit results in additional 155 deaths from road traffic accidents over four months. Then we might view the 155 deaths as a catastrophe, even though the individual deaths are not viewed as catastrophes. And then we should, given Peterson's definition, select a utility function that assigns this collection of deaths from

[6]Instead of defining catastrophe relative to some threshold utility, we could define it relative to some threshold outcome, which would be more apt when a quantitative utility measure is not available.

[7]I am grateful to Martin Peterson for a very helpful discussion of this point.

road traffic accidents a utility below $n$ even though no individual death is assigned a utility below $u$. In contrast, if the 155 deaths are not the result of any single policy intervention, for instance, then it might be more natural to view them as multiple different outcomes. In sum, what counts as a catastrophe, on this view, depends on how we individuate outcomes.

Moreover, suppose that we, as a society, have decided to treat catastrophic risks categorically differently from other risks.[8] Then by the above reasoning, it would follow from Peterson's definition that whether we should respond differently to a risk that is expected to kill say 155 people in multiple incidents than to a risk that is expected to kill the same number of people in one incident, depends on whether we conceive of the former as a risk of a single outcome or as a risk of multiple outcomes.

Another implication of the above way of defining "catastrophe", is that an expected utility maximiser is not, in any significant sense, sensitive to catastrophic risk. An expected utility maximiser is someone who, when confronted with a set of alternatives,[9] always chooses the alternative with greatest expected utility (or one of the alternatives with the greatest expected utility if more than one alternative come out on top). Assuming that the probabilities with which the available alternatives result in different outcomes are known (an assumption that I shall later relax), the expected utility of alternative $A_i$ can be defined as follows.[10] Let $O = \{o_1, ..., o_n\}$ be the set of possible outcomes, and suppose that alternative $A_i$ results in outcome $o_j$ with

[8] To take an example, we might decide to maximise expected utility when managing mundane risks while using some more cautious decision rule (such as one of the cautious rules discussed in section 3) when managing catastrophic risks.

[9] Throughout this article, I will refer to the objects of choice as "alternatives".

[10] The most commonly used version of expected utility theory with known probabilities was developed by John von Neumann and Oskar Morgenstern (1944).

probability $p_{i,j}$. Then the expected utility of $A_i$ is:

$$EU(A_i) = \sum_{j}^{n} u(o_j) \cdot p_{i,j}$$

In other words, the expected utility of an alternative is a probability weighted average of the utilities of the alternative's possible outcomes.

It should be evident that the above equation leaves no room for a special treatment of catastrophes, as defined above. When comparing two alternatives, the only thing that matters, according to an expected utility maximiser, is the alternatives' expected utility. Therefore, the fact that an alternative could result in an outcome whose utility falls below some threshold has no bearing on whether or not it should be chosen, except in so far as it affects the alternative's expected utility. For instance, the fact that an alternative might result in some particular catastrophe does not mean that it will not be chosen, as long as either the catastrophe is sufficiently unlikely or the other outcomes that the alternative might result in are sufficiently desirable.

Since many people think that rationality requires that we maximise expected utility—at least when the relevant utilities and probabilities are well-defined and known—the above might suggest that aversion to catastrophic risk is not rational.[11] Even more worryingly, for those who think that catastrophic risk aversion can be rational, Peterson (2002a) proves that aversion to catastrophe, defined as above, violates some principles that are strictly weaker than the requirements of expected utility theory.

[11]The two most common justifications of expected utility maximisation are that, first, an expected utility maximiser can expect to do better (by her own lights) in the long-run than someone who is not an expected utility maximiser; and, second, that expected utility maximisation is entailed by some purported principles (or axioms) of rational preference (see, e.g., Briggs 2019).

## 2.3 Catastrophe as concentration of fatalities

An alternative definition of catastrophe, that has become quite widespread in theoretical economics and social choice theory, essentially assumes that the decision-maker of interest is an expected utility maximiser. The definition stems from Keeney's (1980) seminal work on attitudes to risk, but it has been extended by for instance Rheinberger and Treich (2017) and Bernard et. al (2018).[12] For the sake of simplicity, I shall focus on Keeney's original definition, but my remarks apply equally to the more recent generalisations.

Strictly speaking, the definition in question does not define "catastrophic" as a category of outcomes, but rather as a category of risks. Informally, the definition says that a small probability of a great number of fatalities is more catastrophic than a greater probability of a smaller number of fatalities, assuming that the expected number of fatalities from the two risks is the same. For instance, a one in a million chance of one million deaths is more catastrophic than one death for sure. If $p$ and $p'$ denote two probability values, $k$ and $k'$ denote two quantities of fatalities, and $(x, y)$ denotes an ordered pair consisting of $x$ and $y$, then Keeney's definition can be more formally stated as follows:

**Definition** (Keeney on catastrophic risk)**.** *Risk* $(p, k)$ *is more catastrophic than risk* $(p', k')$ *if* $pk = p'k'$ *and* $k' < k$.

What does this definition imply for the opening example of this article, for instance, how 155 deaths in a single train accident compare to 155 deaths spread across one hundred road traffic accidents? As previously mentioned,

[12] Other recent work that uses Keeney's or similar definitions include Bommier and Zuber (2008), Fleurbaey (2010), Bovens and Fleurbaey (2012).

the definition does not directly say anything about how to compare these *outcomes*. However, it does tell us how to compare the risk from, say, road traffic with the risk from train traffic. As an illustration, let's make the (obviously false) assumption that we know for sure that exactly one person will die each day in road traffic accidents and we know that there is an independent 1/155 chance each day there will be a fatal train accident (and, for the sake of simplicity, let us assume that if there is a train accident then exactly 155 people will die). Then Keeney's definition, and subsequent generalisations, imply that the risk from road traffic is less catastrophic than the risk from train traffic. And if society should be (or simply is) catastrophe averse, then it should be more concerned with risks from train traffic than from road traffic.

One limitation of Keeney's definition is the assumption of a fixed expectation of fatalities. However, this limitation is arguably justified given Keeney's aims, which was to try to isolate the effect that catastrophe aversion has on the social preference. For if the expected number of fatalities differs between two risks, then a preference for one over the other might be driven by considerations about the expected number of fatalities.

Another apparent limitation of Keeney's definition is that it only explicitly concerns *fatalities*. This limitation is only apparent, however, as "fatality" can be interpreted as the loss of any good. For instance, we could interpret $k$ as the number of insect species lost or as hectares of deforestation. In fact, for Keeney's definition to take into account the *indirect* effects of fatalities on the population in which they occur, $k$ would have to be interpreted very broadly, possibly as broadly as total welfare loss. Some losses of lives, animals or plants are more destabilising to, say, the social or natural order than other

losses. Such a destabilisation is often what makes a loss catastrophic (cf. Boström and Milan Ćirković 2008). Hence, for Keeney's definition to be at all plausible, "fatality" must not be understood too literally.

A more serious limitation of Keeney's definition concerns the fact that it is stated in terms of precise probabilities. Suppose that as a society, we have decided that we should be particularly concerned with catastrophic risks, in that we should treat them categorically differently from more mundane risks (see fn. 8). Now, we cannot use the definition in question to determine whether risks such as those imposed by artificial intelligence and geo-engineering—risks that are typically classified as "catastrophic"—are catastrophic or not, since in those cases we lack knowledge of precise probabilities. So, the decision to give a special significance to catastrophic risk, coupled with Keeney's definition, has no implications for how we should respond to what we would typically consider to be catastrophic risks.

## 2.4 Catastrophe and discontinuity

Yet another suggested definition of "catastrophe" requires some sort of *discontinuity* between catastrophic and non-catastrophic outcomes. Posner (2004: 6), for instance, defines catastrophe as an outcome[13] that, if it obtains, "will produce a harm so great and sudden as to seem discontinuous with the flow of events that preceded it." Note that the previous two definitions do not re-

[13]Strictly speaking, Posner defines catastrophe as a particular type of *event*. In decision theory, "events" are typically understood as collections of "states of the world" that are outside of an agent's control, whereas "outcomes" are taken to be the results of an agent's acts and are thus at least partly under her control. (I thank a reviewer for encouraging me to clarify this.) Given this terminology, a catastrophe could be either an outcome or event, but I shall typically be using the term "outcome".

quires such discontinuity. For instance, on Peterson's definition, there could be an outcome whose utility is exactly $n$, which means that it is catastrophic but would no longer be catastrophic if it were improved even to just the slightest possible degree. Similarly, on Keeney's definition, a single potential fatality might determine how two risks compare in terms of the "more catastrophic than" relation. For instance, even if $(p, k)$ is more catastrophic than $(p', k')$, $(p, k)$ might not be more catastrophic than $(p'', k'')$ which only differs from $(p', k')$ by a single potential fatality. So, given Peterson's and Keeney's definitions, catastrophic outcomes need not be discontinuous with non-catastrophic outcomes.

How does Posner's definition of "catastrophe" handle the opening example of this article? Arguably, the Kaprun train accident was, for Austrians, "discontinuous with the flow of events that preceded it", unlike both each of the individual road traffic accident and the collection of road traffic accidents that resulted in 155 fatalities. Hence, on Posner's definition, the Kaprun train accident might count as a catastrophe while equally many fatalities from several road traffic accidents would not count as a catastrophe. Thus catastrophe averse societies should be willing to do more to prevent events like the Kaprun train accident than to prevent a number of separate road traffic accidents that together would result in an even greater number of fatalities.

It is worth noting that if catastrophic outcomes really are, in some important sense, discontinuous from more normal outcomes, then it might be rational to employ some method other than expected utility theory to decide how to choose when faced with catastrophic risk. After all, the expected utility formula, as defined above, makes no room for such discontinuity. For

instance, no matter how disastrous an outcome would be, an expected utility maximiser is willing to risk ending up with that outcome for the sake of a sufficiently high chance at some outcome that is only marginally better than the status quo.

I shall not in this article take a stand on which of the above (or indeed any other) definition of "catastrophe" is most useful, natural, or correct. Instead, I shall now turn my attention to the different types of situations in which one might be confronted with catastrophic risk, and the different questions that these situations raise about how to respond to such risk.

## 3 Different levels of uncertainty

In the last section it was assumed that the catastrophes had well-defined and known probabilities.[14] In what follows, I shall refer to such catastrophic risk as *quantitative catastrophic risk*. Sometimes, we find ourselves in situations where although we may not know the probability of a catastrophe, we may know how the probability of a catastrophe compares with the probabilities of various other possible outcomes. I call such catastrophic risk *qualitative catastrophic risk*. At other times, we do not even know such comparative probabilities, but are merely aware of the possibility of some catastrophe. But unfortunately, we are sometimes not even aware of some possible catastrophe.

Clearly, quantitative catastrophic risk entails qualitative catastrophic risk which entails awareness of catastrophic risk. Another way to think of this

[14] In other words, it was assumed that the decisions in question are what economists typically call decisions under *risk*. Since we often talk about risks in ordinary language (and in disciplines other than economics) when the probabilities are unknown, and perhaps even undefined, I shall not follow the economists' convention.

categorisation is in terms of the severity of the uncertainty: the uncertainty is most severe when we are not even aware of the catastrophic risk, less severe when we know the comparative probabilities, and even less severe when we know the quantitative probabilities. The appropriate response to a catastrophic risk may depend on the severity of the uncertainty, as the following subsections illustrate.[15]

## 3.1 Quantitative catastrophic risk

The probability that an asteroid will collide with earth can be calculated with good confidence. For all practical purposes, we can treat such a probability as a well-defined and objective probability; roughly, a quantity that, were we to come to know it, we should let it guide our subjective probabilities (cf. Lewis 1980). So, the catastrophic risk imposed by an asteroid is what I call a "quantitative catastrophic risk".

Since the probability of an asteroid colliding with earth is known, standard expected utility theory, as described above, should be applicable to situations where we are trying to decide how to react to a possible collision. However, quantitative catastrophic risk raises some questions that might put pressure on expected utility theory.

One question we can ask about quantitative catastrophic risk is whether we should, when confronted with such risk, accept the type of continuity that expected utility theory entails. For instance, suppose that we are considering increasing the global GDP by 0.1%, which we prefer to the current status

[15]Some argue that the rational response to any risk (not only catastrophic risk) partly depends on the severity of uncertainty about the risk. See e.g. Binmore (2009), Gilboa et al. (2009), and Stefánsson and Bradley (2019).

quo. But now imagine that we learn that doing so brings with it increased risk of human extinction. The continuity that expected utility theory implies ensures that there is some probability $p$ such that we should be willing to start a project that has a probability $p$ of increasing global GDP by 0.1% without human extinction, but also has a probability $1 - p$ of human extinction.

Some may find the above implication of continuity to be hard to accept. However, it should be kept in mind that we all regularly make choices that bring with them some small increases in the risk of our own death for the sake of achieving moderately desirable outcomes. For instance, we might cross the road to pick up a $10 bill we dropped, even if there is some risk that we will be run over by a car in the process.

Another question we can ask about quantitative catastrophic risk is whether the *linearity* assumption of expected utility theory should be accepted when faced with such risk. This linearity assumption for instance implies that decreasing catastrophic risk by 0.01 is always equally valuable, irrespective of the original level of risk. For instance, reducing the risk of a catastrophe from 0.01 to 0 is exactly as valuable as reducing the risk from 0.51 to 0.5.[16] But one might wonder whether that is really the case, in particular, when faced with potential catastrophes (Stefánsson 2020).

Finally, we might ask whether *de minimis* cutoffs are ever justified when confronted with quantitative catastrophic risk.[17] According to standard prac-

[16]Here is an illustration. To simplify, we can assume that the outcome that obtains if the catastrophe is avoided is assigned a utility 0 (but the claim in no way depends on this assumption). And let's consider some particular catastrophe, $C$, that has some (negative) utility, $u(C)$. Then the expected utility of reducing the risk of this catastrophe from 0.51 to 0.5 is $(0.5 - 0.51)u(C) = -0.01u(C)$, which is the same as the the expected utility of reducing the risk of this catastrophe from 0.01 to 0, $(0 - 0.01)u(C) = -0.01u(C)$. (Note that as $u(C)$ is negative, $-0.01u(C)$ is positive.)

[17]This question can also be asked about qualitative catastrophic risk, where *de minimis*

tice in policy and law, risks that are deemed *de minimis*, that is, trivial or minor, are either ignored or treated qualitatively differently than other risks (for discussions, see e.g. Mumpower 1986, Peterson 2002b, Adler 2007). Although it might seem obvious that ideally rational and informed agent's would never employ such cutoffs, it could be argued that it would be reasonable for actual agents to do so, since it might significantly reduce the cost and time of deliberation.

However, if there is no limit to how bad a catastrophe can be, as e.g. Weitzman (2009) assumes, then the reasonableness of *de minimis* cutoffs could be questioned in situations of catastrophic risk.[18] Moreover, even if the badness of catastrophes is limited, one might wonder whether *de minimis* cutoffs can be coherently applied by ordinary agents who lack perfect information, recall and foreknowledge. For instance, even though the risk imposed by each chemical in a set of chemicals is *de minimis*, the risk imposed by the set might not be *de minimis*. So, a regulator who applies *de minimis* when considering which chemicals to allow, might end up ignoring the risks from a set of chemicals even though together the chemicals consistute a risk that is not *de minimis*. More generally, limited agents who apply *de minimis* cutoffs will over time take greater risks than they would upon reflection want to take (Lundgren and Stefánsson 2020).

would be defined relative to some threshold outcome rather than some quantity.

[18]Broome (2013: S30) however points out that this assumption seems questionable in light of the fact that "We are a finite species living on a finite planet".

## 3.2 Qualitative catastrophic risk

Global greenhouse gas emissions impose catastrophic risks that arguably are qualitative. Wagner and Weitzman (2015), for instance, claim that there is *at least* a 10% chance of global warming having catastrophic social and economic consequences. The risk is thus qualitative, in that we know that the probability of a climate catastrophe is at least as great as the probability of randomly drawing a yellow ball from an urn containing nine red balls and one yellow ball.

Sometimes when faced with qualitative risks, we can use *sets* of probabilities, rather than a single probability, to represent our uncertainty. For instance, if the probability of a climate catastrophe is at least 10% but no greater than 20%, then we can represent our uncertainty by a set of probability functions that assign from 0.1 to 0.2 probability to a climate catastrophe. In that case, we might want to use the *Maxmin* expected utility rule (Gilboa and Schmidler 1989), which selects an alternative whose minimum expected utility (relative to all the probability functions in the set that represents our uncertainty) is no lower than other available alternatives' minimum expected utility. Alternatively, since the Maxmin expected utility rule is arguably too cautions, we might instead use an "optimism/pessimism" weighted Maxmin expected utility rule, which weighs each alternative's maximum and minimum expected utility by a factor that reflects the decision-makers optimism/pessimism (see e.g. Binmore 2009).

Note that the above two rules are generalisations of the standard expected utility rule, since they yield the standard rule when the decision-maker only considers one probability function. But situations of qualitative catastrophic

risk might tempt us to use modes of reasoning that depart in more radical ways from expected utility maximisation. For instance, we might ask whether we should employ some version of the Precautionary Principle, which its proponents generally present as a genuine alternative to the rule to maximise expected utility. Although there is no real consensus about how to interpret the principle, it is generally agreed that it implies that in situations of "scientific uncertainty" special precaution should be taken if our actions could result in a catastrophe.[19]

We might also, in situations of qualitative catastrophic risk, consider postponing making a decision until we get better evidence. Now, one might wonder whether postponing isn't just tantamount to choosing one of the available alternatives, in which case postponing should not be seen as an alternative to the decision-rules discussed above.[20] However, one can also see postponing as a meta-decision, that is, one can see it not as a result of a deliberation about a decision-problem after one has formulated it, but rather as a decision to wait with formulating the decision problem until further information becomes available. Finally, if postponing is too costly (or impossible), then we might consider attaching value to flexibility, which might then result in choosing an alternative that is sufficiently flexible such that we can make the optimal choice if further evidence becomes available.[21]

The questions discussed in the last two paragraphs might also arise when we find ourselves in situations where we are aware of a potential catastrophe but neither have quantitative nor qualitative probabilities, which is the type

[19]For a discussion, see e.g. Steele (2006), Peterson (2006), Aven (2010, 2011) Steel (2014), Stefánsson (2019).

[20]I thank a referee for pressing me on this.

[21]Bradley and Steele (2015) discuss some of the issues raised in this paragraph.

of situation to which I turn next.

## 3.3 (Un)awareness of catastrophic risk

The existential risk posed to humanity by artificial intelligence is arguably a catastrophic risk that we are aware of but for which we lack both quantitative and qualitative probabilities. Although experts are often willing to guess the probability of human extinction due to AI, they would readily accept that their guess is nothing more than that, and one might doubt the prudence of basing one's decisions on the exact quantities or comparisons from such guesses.

When one is "merely" aware of catastrophic risk, without knowing either the quantitative or qualitative probabilities, it can be particularly hard to decide what to do. But amongst the principles and strategies that we might consider, is to use the (deterministic) Maxmin rule, which chooses an alternative whose worst possible outcome is no worse than the other alternatives' worst possible outcomes. Alternatively, we might try to postpone the decision until more information becomes available, or choose an alternative that is sufficiently flexible such that an optimal choice can be made if more information becomes available.

The arguably most unfortunate situations, are those where one is not even aware of some potential catastrophe. Situations of unawareness can be distinguished depending on whether the decision-maker realises that there might be some important (perhaps catastrophic) possible outcome that she is unaware of—which is often called *conscious unawareness*—or whether she is even "unaware of her unawareness". The latter type of situation raises no

interesting questions about rational choice: If a person is unaware that her choice might result in some catastrophe, and doesn't even suspect that she is unaware, then it would seem meaningless to ask how she should respond to the potential catastrophe.[22] However, one might ask whether the person was rationally permitted to be so unaware—but that is a question about rational belief, not rational choice.

In contrast, conscious unawareness, or "awareness of unawareness", raises some interesting but difficult questions about rational choice. I think it is fair to say that there is no consensus as to how an agent should rationally choose in situations of conscious unawareness. But the literature on conscious unawareness, and unawareness more generally, is rapidly growing.[23]

# 4 Catastrophe avoidance vs. equity promotion

Before concluding this article, I shall briefly discuss tradeoffs that may arise between on one hand promoting equality and on the other hand reducing catastrophic risk. Let's first consider the goal of promoting *outcome* equality, or *ex post* equality as it is often called. Suppose that we are in a choice situation where the relevant outcomes are that two individuals, Ann and Bob, either live or die. Further, suppose that Ann and Bob are the only two doctors in a remote village that would have great difficulties in attracting new doctors. Now imagine that we have a choice between two risky alternatives:

[22]Here I am assuming the subjective (rather than objective) sense of "should" that is typically of interest to those working on rational choice and risk.

[23]See, for instance, Bradley (2017), for a recent discussion of unawareness and rational choice. Steele and Stefánsson (ta) are a new book-length philosophical treatment of unawareness. Schipper maintains a bibliography on unawareness at http://faculty.econ.ucdavis.edu/faculty/schipper/unaw.htm.

If we choose alternative $A$, then either only Ann dies (with probability 0.5) or only Bob dies (also with probability 0.5), but if we choose alternative $B$, then either both Ann and Bob die (with probability 0.5) or neither dies (also with probability 0.5). This choice situation is represented by table 1.

| | $p = 0.5$ | $p = 0.5$ |
|---|---|---|
| $A$ | Only Ann dies | Only Bob dies |
| $B$ | Ann and Bob die | Nobody dies |

Table 1: Catastrophe avoidance vs. *ex post* equality

To the extent that we are motivated to promote outcome equality, we might be tempted to choose $B$, where Ann and Bob share the same fate, over $A$, where they don't share the same fate. However, it is easy to imagine that $B$ is catastrophic whereas $A$ is not: While only Ann (or only Bob) dying would be merely tragic, it would be catastrophic (for the village) if both doctors die. And, indeed, according to Keeney's definition, $B$ is more catastrophic than $A$: Both alternatives have an expected fatality rate of one, but $B$ has a 0.5 chance of two fatalities wheres $A$ will result in only one person dying. So, in this case, there is a tradeoff between outcome equality and catastrophe avoidance (for a discussion, see e.g. Bovens and Fleurbaey 2012).

Let's now turn our attention to the goal of promoting equality in the distribution of chances, or *ex ante* equality (or risk equity) as it is often called. And imagine now that we have a choice between two slightly different alternatives: If we choose $C$, then Ann will die for sure; but if we choose $D$, then Ann and Bob each have an independent 0.5 risk of dying. (That they have an independent risk of dying means that, say, the probability that Ann dies, given that Bob dies, is the same as the probability that Ann dies give that Bob

survives.) This choice situation is represented by table 2.

| | $p = 0.25$ | $p = 0.25$ | $p = 0.25$ | $p = 0.25$ |
|---|---|---|---|---|
| $C$ | Only Ann dies | Only Ann dies | Only Ann dies | Only Ann dies |
| $D$ | Ann and Bob die | Only Ann dies | Only Bob dies | Nobody dies |

Table 2: Catastrophe avoidance vs. *ex ante* equality

To the extent that we are motivated by *ex ante* equality, we might be tempted to choose $D$ rather than $C$, since $D$ distributes the fatality risk equally amongst Ann and Bob, whereas $C$ imposes all the risk on Ann only. But again, by assumption, $D$ is catastrophic whereas $C$ is not. And, indeed, by Keeney's definition, $D$ is more catastrophic than $C$, since both have an expected fatality rate of one but $D$ has a 0.25 chance of resulting in two fatalities whereas $C$ will result in only one fatality. So, in this case, there is a tradeoff between *ex ante* equality and catastrophe avoidance (Keeney 1980 is an influential discussion of this type of tradeoff).

# 5 Concluding remarks

We have seen that catastrophic risk raises several deep and difficult philosophical questions. For instance, it is far from obvious how "catastrophe" should be defined, nor why (or indeed whether) catastrophic outcomes differ qualitatively from non-catastrophic outcomes. Catastrophic risk moreover raises questions about how to rationally respond, and might be seen as putting pressure on traditional expected utility theory. Finally, catastrophic risk raises hard questions about what to do when the goal to reduce catastrophic risk conflicts with the goal of promoting equality.

## References


Adler, M. D. (2007). Why de minimis? University of Pennsylvania, Institute for Law & Economics, Research Paper no. 07-12.

Aven, T. (2010). *Misconceptions of Risk*. John Wiley & Sons.

Aven, T. (2011). On different types of uncertainties in the context of the precautionary principle. *Risk Analysis*, 31(10):1515–1525.

Bernard, C., Rheinberger, C. M., and Treich, N. (2018). Catastrophe aversion and risk equity in an interdependent world. *Management Science*, 64(10):4490–4504.

Binmore, K. (2009). *Rational Decisions*. Princeton University Press.

Bommier, A. and Zuber, S. (2008). Can preferences for catastrophe avoidance reconcile social discounting with intergenerational equity? *Social Choice and Welfare*, 31(3):415–434.

Boström, N. and Ćirković, M. (2008). Introduction. In Boström, N. and Ćirković, M., editors, *Global Catastrophic Risk*. Oxford University Press.

Bovens, L. and Fleurbaey, M. (2012). Evaluating life or death prospects. *Economics and Philosophy*, 28(2):217–249.

Bradley, R. (2017). *Decision Theory with a Human Face*. Cambridge University Press.

Bradley, R. and Steele, K. (2015). Making climate decisions. *Philosophy Compass*, 10(11):799–810.

Briggs, R. A. (2019). Normative theories of rational choice: Expected utility. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, fall 2019 edition.

Broome, J. (2013). A small chance of disaster. *European Review*, 21(S1):S27–S31.

Fleurbaey, M. (2010). Assessing risky social situations. *Journal of Political Economy*, 118(4):649–680.

Gilboa, I., Postlewaite, A., and Schmeidler, D. (2009). Is it always rational to satisfy Savage's axioms? *Economics and Philosophy*, 25(3):285–296.

Gilboa, I. and Schmeidler, D. (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics*, 18(2):141–153.

Keeney, R. L. (1980). Equity and public risk. *Operations Research*, 28(3):527–534.

Lewis, D. (1980). A subjectivist's guide to objective chance. In Jeffrey, R. C., editor, *Studies in Inductive Logic and Probability*. University of California Press.

Lundgren, B. and Stefánsson, H. O. (2020). Against *de minimis*. *Risk Analysis*, in press.

Martin, I. W. R. and Pindyck, R. S. (2015). Averting catastrophes: The strange economics of scylla and charybdis. *American Economic Review*, 105(10):2947–85.

Mumpower, J. (1986). An analysis of the de minimis strategy for risk management. *Risk Analysis*, 6(4):437–446.

Peterson, M. (2002a). The limits of catastrophe aversion. *Risk Analysis*, 22(3):527–538.

Peterson, M. (2002b). What is a *de Minimis* risk? *Risk Management*, 4(2):47–55.

Peterson, M. (2006). The precautionary principle is incoherent. *Risk Analysis*, 26(3):595–601.

Posner, R. A. (2004). *Catastrophe: Risk and Response*. Oxford University Press.

Rheinberger, C. M. and Treich, N. (2017). Attitudes toward catastrophe. *Environmental & Resource Economics*, 67(3):609–636.

Steel, D. (2014). *Philosophy and the Precautionary Principle: Science, Evidence, and Environmental Policy*. Cambridge University Press.

Steele, K. (2006). The precautionary principle: a new approach to public decision-making? *Law, Probability and Risk*, 5(1):19–31.

Steele, K. and Stefánsson, H. O. (ta.). *Beyond Uncertainty: Reasoning with Unknown Possibilities*. Cambridge University Press.

Stefánsson, H. O. (2019). On the limits of the precautionary principle. *Risk Analysis*, 39(6):1204–1222.

Stefánsson, H. O. (2020). Gambling with death. *Topoi*, 39(2):271–281.

Stefánsson, H. O. and Bradley, R. (2019). What is risk aversion? *British Journal for the Philosophy of Science*, 70(1):77–102.

von Neumann, J. and Morgenstern, O. (2007/1944). *Games and Economic Behavior*. Princeton University Press.

Wagner, G. and Weitzman, M. (2015). *Climate Shock: The Economic Consequences of a Hotter Planet*. Princeton University Press.

Weitzman, M. (2009). On modeling and interpreting the economics of catastrophic climate change. *The Review of Economics and Statistics*, 91(1):1–19.