The Black Hole Challenge: Precaution, Existential Risks and the Problem of Knowledge Gaps

Christian Munthe

Submit your article to this journal ⏎

View Crossmark data ⏎

Routledge
Taylor & Francis Group

The Black Hole Challenge: Precaution, Existential Risks and the Problem of Knowledge Gaps

Christian Munthe

Department of philosophy, linguistics and theory of science, University of Gothenburg, Gothenburg, Sweden

ABSTRACT

So-called 'existential risks' present virtually unlimited reasons for probing them and responses to them further. The ensuing normative pull to respond to such risks thus seems to present us with reasons to abandon all other projects and commit all time, efforts and resources to the management of each existential risk scenario. Advocates of the urgency of attending to existential risk use arguments that seem to lead to this paradoxical result, while they often hold out a wish to avoid it. This creates the 'black hole challenge': how may an ethical theory that recognizes the urgency of existential risks justify a limit to how much time and resources are committed to addressing them? This article presents two pathways to this effect by appealing to reasons for limiting the 'price of precaution' paid in order to manage risks. The suggestions are different in that one presents ideal theoretical reasons based on an ethical theory of risk, while the other employs pragmatic reasons to modify the application of ideal theoretical ideas. The latter of these ideas is found to be slightly more promising than the first.

1. Introduction

In this article, I will consider a profound challenge to any type of precautionary approach to decision-making. This challenge arises independently of whether one's preferred idea of precaution or a precautionary principle, contrasts itself against standard decision theoretical and risk analytical models or if it holds itself out as a compatible complement to these. The challenge, moreover, arises regardless of what 'degree' or 'price' of precaution such a precautionary idea holds out as defensible and/or required.[1] The challenge has as its root the situation when a decision-maker faces, what I will term, a *knowledge gap*[2]; i.e. a situation where a decision needs to be made, but where there is lack of knowledge regarding relevant facts for assessing what decision should be made. This lack may be more or less profound and may consist of varying degrees of uncertainty or imprecision of existing information, known areas of ignorance, as well as the ever-present possibility of yet unidentified such areas and uncertainties. The extent and nature of such a knowledge gap is always relative to what is considered to be *relevant* facts in light of whatever normative criteria are used for assessing the outcome. For

instance, if the criteria evaluate outcomes in terms of imprecise standards – e.g. intervals of some value, such as ranges of life-expectancy or number of deaths – certified knowledge about outcomes in such imprecise terms would not contribute to a knowledge gap. But if the assessment criteria are more precise – e.g. requiring an exact number of life-years gained or lost, or the number of deaths caused or prevented – imprecision, however, certified, will create a knowledge gap in the form of lack of knowledge of what a precise outcome would be.

All real-life decisions have the decision-maker face some kind of knowledge gap. Therefore, an idea of precautionary decision-making needs to be able to guide decision-makers with regard to:

(1) if the knowledge gap faced is to be tolerated, and a decision made in spite of it, or
(2) if the decision should be delayed while attempting to close or narrow the gap,
(3) and, if so, how much time, effort and resources should be spent on that endeavor.

This 'problem of knowledge gaps' is a well-known theme across many types of decision-making, and recently linked to debates about precautionary decision-making through Steel's notion of 'epistemic precaution' (Steel, 2014, ch. 8). It has been recognized to present fundamental challenges to standard decision theoretical models (Levi, 1986; Sobel, 1994), augmented when the ambition is to underpin a normatively valid approach to precautionary decision-making (Munthe, 2011).[3] In practice, the problem of knowledge gaps may often be pragmatically overcome by institutional solutions, e.g. as procedural rules within legal systems (Laudan, 2008), or as set evidence type criteria in the regulatory systems for licensing drugs or toxic substances (see, e.g. Steel, 2014, ch. 9). As the importance of coming to a decision (in a particular case or a certain type of cases) outweighs the importance of further certainty at *some* point, a decision-making system with institutional routines that produce closure without justifying why the call has to be made precisely at that point can be accepted as good enough (Munthe, 2011, ch. 6).

However, this type of solution is not as obviously available for the challenge I wish to address in the present context. This challenge results when the problem of knowledge gaps is combined with the presence of what I for convenience sake will term an *existential risk*. This term, launched in recent discussions about the ethics of science and technology,[4] indicates not a risk in the technical sense of a precise probability of a possible future event combined with a precise outcome value of that event, but exactly a knowledge gap, where this includes *a priori* possibility or vague likelihood of an extremely bad (often irreversible or irreparable) outcome, what has been termed 'ultimate harm' (Persson & Savulescu, 2012). Admittedly, there is a tendency in this field to be rather anthropocentric and disregard scenarios other than those entailing the end of humanity. This, however, says more about author bias than about the theoretical notions of an existential risk and ultimate harm as such. The latter may be formulated and envisioned on the basis of any conception of the good, e.g. a bio- or ecocentric conception, as long as it allows for scenarios where everything or almost everything that possess such goodness is destroyed or otherwise made to vanish. Based on any such underlying theory of value, we need not make very strong ethical assumptions to deduce extraordinary moral reasons to avoid ultimate harms. These reasons, in turn, feed into our reasons to manage existential risks – the massiveness of the threat makes it important

to mind quite a bit also about distant possibilities and very vague and uncertain likelihoods, and at the very least make efforts to clarify them and how they might be managed.

Existential risks basically come in two versions (Häggström, 2016): either a threat of this sort is at hand independently of human action, and a decision-maker needs to ponder how to respond to it or it arises due to otherwise valuable human action,[5] making the decision-making problem about whether or not to tolerate the threat and, if not, how to modify alternative options in the face of the threat. The first version is exemplified, e.g. by the threat of a large meteorite crashing into Earth, while the second is illustrated by the possibility of (otherwise beneficial) bio- or nanotechnological inventions running amuck in nature. Facing the first kind of situation, a decision-maker needs to decide how much time and resources to spend on managing the threat (in which case, a problem of knowledge gap with regard to the effectiveness of contemplated measures, and the importance of contemplating these measures, will be faced). In the second instance, the decision-maker instead needs to decide how much resources and time should be spent on clarifying the risk of the technology before making a decision on its use.

The presence of an existential risk implies that the apparent plausibility of the sort of pragmatic, institutional solution to the problem of knowledge gap mentioned above withers. Accepting the idea that ultimate harms justify extraordinary reasons for precautionary management of the knowledge gaps presented by existential risks, there seems to be no rationally justifiable end to the amount of resources and time that we should spend on trying to clarify an existential risk further. This is due to the very nature of the harms of being ultimate: this fact seemingly cancels any claim to the effect that making a decision with regard to it may be more important than making sure about the harm itself. Each such risk, therefore, appears as a 'black hole' that, by the awesome normative power of the mere ultimateness of the harm it threatens us with, sucks into its void all other things that might have been mobilized for the pursuit of the good in human decision-making. The more precise ingredients in this 'black hole challenge', and what makes it into such a difficult problem will be discussed at length in the next section.

In this article, I will sketch two possible paths for responding to the black hole challenge while keeping the normative commitment to a precautionary reason for managing knowledge gaps responsibly. Both paths rest on former attempts to develop a moral theory of precaution based on an ethics of risk but differ in that they force potentially conflicting aspects of such theories to the disadvantage of the other. Before the sections where these proposals are set forth, I will make some preliminary remarks and clarifications with regard to the nature and difficulty of the problem, addressing a few potential objections underway.

2. Preliminaries: Proportionality, the Price of Precaution and the Real Lesson of Pascal's Wager

Three insights have gradually grown across all who have contributed to the philosophy and ethics of precaution (Munthe, 2013, 2015a, 2016). First of all, as indicated, due to the need of handling the problem of knowledge gaps, a sound notion of precaution in decision-making must transcend standard models of risk-cost-benefit analysis. Second, such a notion also needs to retain certain core elements of these standard models, not least requirements to consider all options including *status quo*, and requirements of

considering decision and opportunity costs. This is necessary in order to avoid different sorts of precautionary paradoxes, such as what Steel has called 'inconsistency' (Steel, 2014), and what Munthe calls 'decisional paralysis' (Munthe, 2011). Third, normative justification of particular notions of required precaution needs to include some idea of what Steel has called *proportionality* (Steel, 2014), and this requirement may itself be explained by the concept of the 'price of precaution' (Munthe, 2011, ch. 1).

This price indicates the need for any idea of precaution in decision-making to specify how much in terms of direct, indirect and opportunity costs should be acceptable in order to enact some degree of precautionary action (Munthe, 2011, ch. 1). This, however, implies a need for normative justification that goes beyond Steel's own proposals of 'efficiency' (John, 2016; Munthe, 2015a, Steel, 2014). This proposal requires that the costs of enacting a given degree of precaution should always be minimized, but leaves completely open *what* level of precaution should be aimed at. Therefore, it remains undecided among a wide range of proposals on how much of 'precaution for the buck' is required for a precautionary policy to be justified. Because of that, this proposal needs to be complemented to face up to the black hole challenge. For consider, once again, the awesome gravity of an existential risk: once we have faced up to responding to it, we have signed up for the prospect of having every last cent being pulled into its black hole even if we accept a cost minimizing requirement.

Thus, demonstrating what a proportional price of precaution in the face of existential risk amounts to requires more than Steel has provided, and what it requires needs to be of a normative ethical nature. It is far from clear, however, what exactly is needed and how it may be justified. A recent risk ethical theory pursuing the problem of knowledge gaps (Munthe, 2011) does, for instance, strike me as too unspecific, and too much dependent on pragmatic patching at the institutional policy level to do the trick, and similar things hold for approaches based on political (Gardiner, 2006; McKinnon, 2012) and moral philosophy (Hansson, 2013). As mentioned, institutional pragmatic solutions to the lack of specific normative decision guidance hold up to rational scrutiny as long as we consider more ordinary decision problems, but falters in the face of existential risks. At the same time, the core aspects from orthodox decision theory and risk-cost-benefit analysis that normally help us handle also very complex risk decisions and precautionary policy challenges seem just as feeble in the face of the black hole challenge.

To press this last point, I will close this section by revisiting a debate coming out of two blog posts from last year (Munthe, 2015b, 2015c). In these, the rationale behind strong advocacy of attending to existential risks was held out to resemble that of *Pascal's wager* (Hàjek, 2012). In effect, all of these advocates either owe us an explanation for where to stop ensuring against the eventual end of humanity/life on earth/ sentient life in the universe or for why they are not advocating the attendance of mass in accordance with the famous Pascalian prescription for ensuring against possible eternal torment in hell. The one published response to this, by Olle Häggström (2016, pp. 242–245), has tried to deflect the challenge in three ways. First, by attacking Pascal's wager itself with the familiar observation that, as there exist so many incompatible teachings on what one must do to escape eternal damnation, the wager does not support any particular course of action. However, this argument seems to hit existential risk response advocacy just as much. As there are so many different ways in which we might imagine nature or human action to effect ultimate harms, the existential risk

argument will not support much of specific action in the first place (otherwise we would have a precautionary paradox). Or, alternatively, it will support putting all our monies into managing the first existential risk we happen to come to think about (by the logic of the black hole challenge), but then the advocate of that would seem to have all the reason in the world to step into the first church[6] happening to appear in his or her path and remain there praying, just to be sure. Note that the principles of decision theoretical orthodoxy mentioned above seem unable to rescue us out of this spot, as each option will motivate spending all our resources on it, and once we have spent them on it, the other ones have seized to be options (as we have no resources to perform them).

Second, Häggström tries to wiggle himself out of the Pascal's wager *analogy* by complaining about the epistemic quality of the likelihoods underlying the risk of that particular argument. The likelihoods of various Abrahamitic religious teachings are too poorly substantiated, while those underlying the existential risk scenarios are not. This is the so-called *de minimis risk* solution to resist precautionary paradox, attempted by many in many variations across debates on the philosophy of risk and precaution. But that move provides no rationale by itself, it simply *assumes* an arbitrarily chosen threshold, while there is no end to suggestions and counter-suggestions on how such a proposal might or might not be justified (Munthe, 2011, 2013; Peterson, 2002; Sandin, 2005). What is more, the idea that we should discount formidable dangers just because they are uncertain seems to run diametrically contrary to the main reason behind the existential risk urgency, namely the massiveness of the ultimate harm implied by their (very uncertain) actualization. Accordingly, in debates on the philosophy of precaution, it is more common to propose limits to what we need to consider in terms conditions about the outcome-dimension of risk being *serious enough*, such as an irreversible catastrophe (Allhoff, 2009; Manson, 2002; Munthe, 2011, 2013). Such conditions surely limit what a precautionary approach to decision-making tells us to consider, but they do not exclude either existential risks or eternal torment in hell. For sure, one may want to propose a condition regarding what quality of evidence is needed in order for some threat to be worthy of precautionary consideration. However, the only ground for calibrating such a condition would seem to be in terms of how much resources and time we should be required to spend on clarifying eventualities, that is exactly the question underlying the black hole challenge. We are back to square one.

Third, Häggström states that he, and, he believes, other advocates of the urgency of attending to existential risks are quite modest and only advocate some moderate probing of existential risk scenarios (thus not qualifying for the Pascal's wager analogy and escaping the black hole challenge). As Bostrom has suggested that attending to existential risk should be a 'global priority' (Bostrom, 2013), I am uncertain whether this characterization is actually true. But suppose that it were, then the challenge simply moves to being about how to justify such modesty in light of the rationale underlying the advocacy of the urgency of existential risks. Or, alternatively, how to justify the commitment to urgently attending to existential risk in light of that modesty. The normative force of possible ultimate harms, once acknowledged in the way it is in the existential risk context, hardly inspires modesty, but rather activates the precise gravity pulling us towards the black hole.

3. Escaping Gravity, Suggestion 1

In a recent theory of the ethics of risk in the context of justifying a principle of precaution (Munthe, 2011, ch. 5), some fundamental details were left hanging, the clarification of which might help to avoid the black hole challenge. This theory rests on, first, the idea that the issue of how much resources and time to spend on clarifying knowledge gaps is a decision problem to which a normative theory of precaution and risk applies. This theory links the idea that a precautionary principle should be viewed as 'guiding beliefs' to the idea that it should be 'guiding action' (Peterson, 2007), abandoning the need to distinguish these understandings: it guides belief by saying what actions should be taken to justify beliefs. So, facing the issue of using a technology T or not, where T (and, by implication, not-T) has us facing a knowledge gap, our decision problem will always be richer, also entailing a number of options to delay the use of T (and the final decision whether or not to use it at all, and how) while updating our information about T and not-T (implying more or less spending of resources and time for that purpose). Assuming that T brings some sort of benefit (or chance of such),[7] such delays will always carry a price of precaution in terms of direct costs for the probing, indirect costs in terms of risks created by actions undertaken to probe and opportunity costs in terms of delaying chances of benefits. To justify that this price is increased through actions meant to narrow or close a knowledge gap, there has to be a moral reason for deciding on more updated information rather than less, that should be factored into the question of how high the price of precaution should be allowed to be. This reason may be supported on different grounds. While the mentioned theory assigns a fundamental moral reason to this effect in terms of what it takes to act responsibly, Steel's idea of epistemic precaution cites instrumental and historical evidence for why we should care about the quality of evidence underlying assessments of risks and chances in practical decision-making. A third option might be to point to broadly conceived virtue-ethical notions of what it takes to avoid morally irresponsible negligent and reckless behavior (Knutsson & Munthe, 2017; Sandin, 2009). For instance, facing a new agricultural biotechnology with uncertain ecological effects but clear benefits, e.g. reducing the use of pesticides or fertilizers, there will be reasons to pay a price in terms of tolerating higher pesticide and fertilizer use for some time while spending resources on investigating the long-term ecological effects of using the technology. To avoid precautionary paradox, this reason cannot be conclusive, it has to be balanced against the reasons provided by the chances of reducing pesticide and fertilizer use. But these latter reasons cannot be conclusive either, they all need to be balanced in some way against each other (Munthe, 2011, ch. 2 & 5).

But *how* should they be so balanced, and on what basis? Steel's theory, as mentioned, leaves that issue open (albeit requiring *some* balance to be struck, and costs to be minimized when effecting that balance). Munthe has suggested two more specific normative ethical solutions, where more importance is put on avoiding risks than securing benefits, while the options and stakes of a situation can make the defensible price of ensuring epistemic precaution very different, in order to avoid precautionary paradox. One of these is a simple idea of a progressive increase of the importance of avoiding risks with larger harm components – the larger such a harm component, the higher the acceptable price of precaution becomes of clarifying that risk in order to secure against the possible

harm (Munthe, 2011, pp. 115–118). This solution would enhance rather than mitigate the black hole challenge, as the massiveness of the ultimate harm of an existential risk will here be bestowed with even more pronounced importance. The other, however, presents a more complex idea, where the presence of a 'good enough' option drastically increases the defensible price of precaution for all options bringing more drastic risk than this one. This variant leaves open a possibility of resisting the black hole challenge, and below this possibility will be described in some detail.

Suppose that the decision problem with regard to T is worked out to contain the options O_0–O_9, where the first and the last are the options to use T immediately, and to never use T, respectively. The other options are all about spending increasing amounts of time and resources to clarify the possible ecological threat posed by T, and ways of managing it. Now, suppose that one of these options presents a 'good enough' solution in terms of the values at stake in this decision problem (including the value of making well-founded decisions), say O_5. Then, O_1–O_4 all imply different degrees of irresponsible lack of consideration to the knowledge gap posed by T in view of the values at stake, and also that the ecological risks left unattended are much less easy to justify than if O_5 (or any other 'good enough' option) had not been available. O_6-O_9 instead imply increasingly overly elevated prices of precaution, increasingly difficult to justify in view of the presence of the 'good enough' O_5. This makes it much more difficult to justify any of the other options, while O_5 entails the idea of at some point making the call on whether or not to use T at all based on the updated basis of information obtained through O_5. In the case of the new agricultural biotechnology, this would mean spending a fair amount of time and resources on clarifying its knowledge gaps, but then make the call on whether or not to use it based on the information available at that time, knowing that even more and more certain information could have been obtained if further time and resources had been used to that effect. An implication of this theory is that what kind of transitions of practices this approach may allow may be relative to what position we are occupying. If T is a new treatment for a serious condition for which there is no existing therapy changes the situation compared to if T is a new treatment for a serious condition where there is an existing decent therapy. In the first case, the acceptable price of precaution is quite low (although it exists, as the treatment may have side-effects that are even worse than the condition), while it increases in the second case due to the presence of a decent alternative while evidence with regard to T is assembled. The trick in all of these cases, as in this model generally, is, of course, to determine what makes an option 'good enough', and that issue has not been solved in presentations of the theory. At the same time, it is the relativization of the acceptable price of precaution to such a 'good enough' option that opens a door to escaping the black hole challenge. If, in the face of an existential risk, we can identify such an option, we are justified to discount the importance of clarifying this risk and/or possible responses to it beyond the price of precaution implied by this option.

Addressing the issue of how to proceed to flesh out the theory with clear conditions for the 'good enough' option, two distinct strategies have been identified (Munthe, 2011, pp. 126–129). One of these is to proceed on the course set by the relativization to the conditions, options and stakes of a particular decision context, and also relativize what may be a 'good enough' option. This strategy entails the idea that every decision

context will have such an option, and that it will – like O_5 above – always be a sort of compromise option, relative to the structure of stakes and knowledge gaps of this particular context. However, this idea seems counterintuitive in several ways. We may imagine decision contexts where all options are more than 'good enough', such as when humanity has increased its life-expectancy to 300 years (evenly distributed), and T offers uncertain prospects (with some unclear risks) to extend it further to 400 or 500 years. In such a situation, a rather elevated price of precaution seems proper for all options, as almost no additional risk seems worth attempting to change an already more than splendid situation. Similarly, we may envisage decision contexts where all options are *worse* than 'good enough', e.g. if we change the scenario to one where the life expectancies are 30, 40 and 50 (Munthe, 2011, p. 128). To these reasons against the relativist solution for determining the 'good enough' option, we may now add the existential risk scenarios in view of the black hole challenge. The idea of a context-dependent compromise solution seems to assume the notion of a middle ground within the range of stakes and uncertainties making up the decision context, but this idea becomes moot when one option seems to present (almost) infinitely powerful reasons for certain courses of actions. That is, if the 'good enough' option is to be determined relative to the range of stakes and uncertainties, and we accept the massive basic normative pull to attend to existential risks, then the black hole challenge pulls us towards the conclusion that the only 'good enough' solution is to put all we have on black and see it disappear into the hole. Accepting the awesome need of humanity to accomplish interstellar exodus to escape meteor catastrophes and such, no other reason seems powerful enough to resist the call for having all our resources put to use to that (quite possibly unattainable) end.

Therefore, to move things ahead, the idea of absolute determinants of what makes for a 'good enough' option relative to which we may determine what price of precaution is acceptable looks more attractive. With such an idea at our disposal, we could say that even in the face of the awesome normative pull of existential risks, there will be a point at which the price of precaution of further clarification and management of these risks becomes drastically more difficult to justify. Despite the massiveness of ultimate harms, there is a limit to how much we should pay, mess up and give up in order to insure against them. We do not need to close down all other production and public services or abandon the management of more clearly identified and manageable large risks, such as the lack of access to clean drinking water or the effective prevention of further antibiotic resistance to secure functioning health-care systems, in order to put all we have into the eventual possibility of future interstellar human migration to insure against meteorite induced extinction or some such. The challenge still remains of determining what features determine the point at which our evidential situation is 'good enough' to make the call, but at least this type of strategy seems *capable* of producing some sort of limit that resists the black hole challenge.

4. Escaping Gravity: Suggestion 2

Suggestion 1 has a clear downside in that it depends on yet to be developed theoretical details. When such development is attempted, it may transpire that the ordered goods are in fact not available for delivery. An alternative strategy, therefore, is to follow

another path from previous work on the ethics of risk and precaution, and complement abstract theory with pragmatic solutions (Munthe, 2011, ch. 6). These are solutions that have a rationale in terms of how they help an institutional system implementing the abstract theoretical aspects of responsible decision-making by complementing these aspects with practical solutions, whose justifiability is not 'provable' by proceeding from the axioms of the abstract theory. Such pragmatic solutions would seem to fit the idea of Per Sandin, Martin Peterson and David Resnik of viewing the precautionary principle as non-ideal theoretical or 'mid-level' (Resnik, 2012; Sandin & Peterson, 2019). In this section, I will very briefly sketch one variant of how such a solution that would help us resist the black hole challenge might look like.

The basis of pragmatics is the nature of human beings and human societies. We may know that a certain solution to a societal problem would be the best one from the view of science, technology and ideal morality, but we may as well know that it would be a catastrophe to attempt implementing it, because it would be resisted or mismanaged by people. This is, in fact, a quite common situation when implementing scientific or idealistic suggestions in political reality. So, the ideal theoretical suggestion needs to be tweaked in order to be implementable, for instance by being complemented or adjusted to become easier to manage in institutional routines or to become easier to accept by the people affected by it. The trick is to be able to accomplish this tweak without losing the qualities providing the reason for the suggestion in the first place. For instance, technical security systems need to adapt to the various limitations of human psychology (demanding very complicated and unique passwords tend to make us write them down in numerous places to remind ourselves, thereby actually creating less security than what would have been accomplished with less complicated and variable solutions), but not too much, as the system then becomes impotent to address the risks it is supposed to prevent being actualized (e.g. using 'qwerty12345' for all passwords).

Moving back into the area of the present article, we may start by noting that pragmatic considerations add reasons against the theoretical suggestion of the former section, as it, even if attainable, is bound to be controversial.[8]

But what aspects of human nature and human societies may be used positively, to pragmatically motivate a limit that may help us resist the black hole challenge? I propose two such aspects: First, a tendency of people to be more motivated to start spending resources and time on ambitious and time-consuming projects where there are concrete action plans, similar to what Steel has termed 'sequential plans' (Steel, 2014, ch. 6), with pre-set conditional end- and exit-points. Second, a tendency of people to resist ever *concluding* ambitious and time-consuming projects, once ventured on, unless there are pre-set end- and exit-points. The first tendency comes out of the psychological mechanism of resisting to commit one's efforts to vague prospects – these will not appear as making the investment worth it. The second tendency instead comes out of a mechanism that once we have committed to a project, we are likely to find new reasons for extending our commitment. I will not here go into the many finer psychological details involved in these aspects of being human, but trust that readers will have no problems identifying several cases in their surroundings providing illustration.

If we now apply this pragmatic aspect to the case of existential risk and the black hole challenge, what transpires is a mirror image to the picture painted in the preceding

section. When the black hole challenge seen from a theoretical standpoint threatens to have the normative pull of ultimate harms suck all available resources into endless endeavors to clarify ridiculous eventualities of horrific things, from a pragmatic standpoint, it instead is a threat against having any willingness to spend even the tiniest bit of time and effort on clarifying existential risks. The awesomeness of one threat of ultimate harm after another silences human ability to psychologically engage with the challenge; venturing into black holes is not an idea that entices us, it either paralyzes us or makes us act irrationally.[9] Likewise, if we engage in precautionary response to a risk that is not existential, but quite substantial and containing challenging knowledge gaps, we will be likely to continue and expanding that engagement as if it had been an existential risk, unless we are guided by clearly stated end- and exit-points. Thus, without end- and exit-points we are likely to (a) never rationally engage in existential risks in spite of the very good reasons to do so, and (b) be consumed by less serious risks beyond what is ethically and rationally motivated. The same points can easily be made at the social level, in terms of what social arrangements we would be likely to accept or to dissolve once accepted. In consequence, from a pragmatic perspective, if we embrace the normative pull of ultimate harms and the strong reason for precaution in the face of existential risks it produces, we should formulate pre-set limits to our precautionary engagement with these risks that will become practical blockers of the black hole challenge.

Also, this strategy leaves the question of what exactly the end- and exit-points should be, but here the prospect of finding ways of determining them seems slightly less uncertain than the one of theoretically justifying what makes for a 'good enough' option. As the reason for the importance of formulating the end- and exit-points now is pragmatic – to make room for sensible precautionary response to existential risks within the limitations of human nature – and not ideal-theoretical, we may allow for some arbitrariness. The important thing is to have the end- and exit-points *reasonably well* placed to do the job *reasonably well*, not to have them perfectly placed to guarantee having the job done nothing short of excellently.

5. Conclusion

The acknowledgment of normative reasons for precautionary responses to knowledge gaps and the normative pull of existential risks, produce what I have called the black hole challenge to any attempt at delivering a plausible philosophical theory of precaution and the ethics of risk. In this article, I have explained how this challenge remains in the face of some standard attempts at dealing with knowledge gaps in decision theory and risk analysis, and some preliminary objections from advocates of the urgency of responding to existential risks. Based on previous work on the philosophy of precaution and the ethics of risk, I have then explored two pathways – one ideal-theoretical and one pragmatic – for justifying avoidance of the challenge without losing the strong normative commitment to respond to existential risks. Both paths leave important details open, but the pragmatic one may do so in slightly less problematic way. The proposal then is that, in spite of the awesome normative pull of ultimate harms, we should let our responses to existential risks be tempered by pre-set limits in order to account for facts of human nature that would otherwise lead us to either fail to act responsibly or accepting irresponsibly elevated prices of precaution.

Notes

1. These words indicate the nowadays common notion that a plausible idea of precautionary decision-making or of the ethics of risk in general, needs to be gradual and scalar rather than absolute or binary. See, e.g. Hansson (2013), Munthe (2011), and Steel (2014).
2. This term is inspired by a terminology of 'gap of knowledge', 'evidence gap', and 'research gap' that has become standard in the literature on evidence assessment or evidence basing of practices, especially so-called Health Technology Assessment.
3. The most recognized approach to the problem of knowledge gaps in decision and economic theory, the so-called decision-cost or value of information idea, is mostly normatively inert or arbitrary. In its philosophically most developed versions, such as in the work of Per-Erik Malmnäs (1994), Malmnäs (1999), it may declare some gaps to be unproblematic as closing them would make no difference, but leave most of them unresolved (Munthe, 2011).
4. See, e.g. Bostrom (2013), Bostrom and Cirkovic (2011), Häggström (2016).
5. The qualification of 'otherwise valuable' is due to the plausible condition that if an option would bring no benefits, it is irresponsible to embark on it (as that would mean creating risks for no good reason at all) (Munthe, 2011, ch. 5).
6. I use this word as denoting any building of worship within any religion holding out rescue from eternal damnation and torment as one of its goods.
7. As mentioned, if that assumption is dropped, there is no reason to use T in the first place, and, since T brings risks, using it will be irresponsible.
8. One may ask if this aspect could not be built into the theoretical solution – the theory of the ethics of risk and precaution – itself, having this theory adjust for the various implementation problems that can be foreseen. This would amount to giving up the distinction between (ideal) theory and (non-ideal) pragmatics. This article is not the place to attempt such a grand project.
9. It may appear that the existence of (a few) enthusiastic existential risk response advocates may speak against this image of human psychology. It does not. Such advocates tend to irrationally spend disproportionate amounts of time and energy on some arbitrarily selected existential risk(s), while they keep having trouble convincing others of the righteousness of their cause.

Funding

References

Allhoff, F. (2009). Risk, precaution, and emerging technologies. *Studies in Ethics, Law, and Technology*, *3*(2). doi:10.2202/1941-6008.1078. Online, Retrieved from https://www.degruyter.com/view/j/selt.2009.3.2/selt.2009.3.2.1078/selt.2009.3.2.1078.xml

Bostrom, N., & Circkovic, M. (Eds.) (2011). *Global catastrophic risk*. Oxford: Oxford University Press.

Bostrom, N. (2013). Existential risk prevention as a global priority. *Global Policy*, *4*, 15–31.

Gardiner, S. M. (2006). A core precautionary principle. *The Journal of Political Philosophy*, *14*(1), 33–60.

Häggström, O. (2016). *Here be dragons: Science, technology and the future of humanity*. Oxford: Oxford University Press.

Hájek, A. (2012). Pacal's Wager. In: E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy* Online, Retrieved from http://plato.stanford.edu/archives/win2012/entries/pascal-wager/

Hansson, S. O. (2013). *The ethics of risk: Ethical analysis in an uncertain world*. London: Palgrave Macmillan.

John, S. (2016). Philosophy and the precautionary principle: Science, evidence, and environmental policy Daniel Steel. *Journal of Applied Philosophy*, *33*(2), 217–218.

Knutsson, S., & Munthe, C. (2017). A virtue of precaution regarding the moral status of animals with uncertain sentience. *Journal of Agricultural and Environmental Ethics*, *30*, 213–224. online first

Laudan, L. (2008). *Truth, error, and criminal law*. Cambridge: Cambridge University Press.

Levi, I. (1986). *Hard choices: Decision Making under unresolved conflict*. Cambridge: Cambridge University Press.

Malmnäs, P.-E. (1994). Towards a mechanization of real-life decisions. In: D. Prawitz & D. Westerståhl (Eds.), *Logic and philosophy of science in Uppsala* (pp. 231–243). Dordrecht: Kluwer Academic Publishers.

Malmnäs, P.-E. (1999). *Foundations of applicable decision theory*. Stockholm: Department of Philosophy, Stockholm University.

Manson, N. (2002). Formulating the precautionary principle. *Environmental Ethics*, *24*, 263–274.

McKinnon, C. (2012). *Climate change and future justice: Precaution, compensation, and triage*. New York, NY: Routledge.

Munthe, C. (2011). *The price of precaution and the ethics of risk*. Dordrecht: Springer.

Munthe, C. (2013). Precautionary principle. In: H. LaFolette (Ed.), *International encyclopedia of ethics* (pp. 4031–4039). Chichester: Wiley.

Munthe, C. (2015a). Precaution, bioethics and normative justification. *Monash Bioethics Review*, *33* (2), 219–225.

Munthe, C. (2015b, February 1). Why aren't existential risk/ultimate harm argument advocates all attending mass? *Philosophical Comment*, Online, Retrieved from http://philosophicalcomment. blogspot.se/2015/02/why-arent-existential-risk-ultimate.html

Munthe, C. (2015c, February 6). An addendum re existential risk arguments: A comment and a fresh application at cern with hawking and de grass tyson at the centre. *Philosophical Comment*, Online, Retrieved from http://philosophicalcomment.blogspot.se/2015/02/an-addendum-re-existential-risk.html

Munthe, C. (2016). Precautionary Principle. In: H. Ten Have (Ed.), *Encyclopedia of global bioethics* (pp. 2257–2265). Cham: Springer.

Persson, I., & Savulescu, J. (2012). *Unfit for the future: The need for moral enhancement*. Oxford: Oxford University Press.

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

Peterson, M. (2007). Should the precautionary principle guide our actions or our beliefs? *Journal of Medical Ethics*, *33*(1), 5–10.

Resnik, D. B. (2012). *Environmental health ethics*. Cambridge: Cambridge University Press.

Sandin, P. (2005). Naturalness and *De Minimis* Risk. *Environmental Ethics*, *27*(2), 191–200.

Sandin, P. (2009). A new virtue-based understanding of the precautionary principle. In: M. A. Bedau & E. C. Parke (Eds.), *The ethics of protocells: Moral and social implications of creating life in the laboratory* (pp. 89–104). Cambridge, MA: MIT Press.

Sandin, P., & Peterson, M. (2019). Is The precautionary principle a midlevel moral principle? *Ethics, Policy & Evironment*, *22*(1). doi:10.1080/21550085.2019.1581417.

Sobel, J. H. (1994). *Taking chances: Essays on rational choice*. Cambridge: Cambridge University Press.

Steel, D. (2014). Philosophy and the precautionary principle: Science, evidence, and environmental policy. Cambridge: Cambridge University Press.