Risk and Resilience For Unknown, Unquantifiable, Systemic, and Unlikely/Catastrophic Threats

Seth D. Baum
Global Catastrophic Risk Institute
http://sethbaum.com * http://gcri.org

Environment, Systems, and Decisions 35(2): 229-236. This version dated 17 June 2015.

Abstract
Risk and resilience are important paradigms for analyzing and guiding decisions about uncertain threats. Resilience has sometimes been favored for threats that are unknown, unquantifiable, systemic, and unlikely/catastrophic. This paper addresses the suitability of each paradigm for such threats, finding that they are comparably suitable. Threats are rarely completely unknown or unquantifiable; what limited information is typically available enables the use of both paradigms. Either paradigm can in practice mishandle systemic or unlikely/catastrophic threats, but this is inadequate implementation of the paradigms, not inadequacy of the paradigms themselves. Three examples are described: (i) Venice in the Black Death plague, (ii) artificial intelligence (AI), (iii) extraterrestrials. The Venice example suggests effectiveness for each paradigm for certain unknown, unquantifiable, systemic, and unlikely/catastrophic threats. The AI and extraterrestrials examples suggest how increasing resilience may be less effective, and reducing threat probability may be more effective, for certain threats that are significantly unknown, unquantifiable, and unlikely/catastrophic.

Keywords: risk; resilience; catastrophe; plague; uncertainty; artificial intelligence; extraterrestrials

1. Introduction

Risk and resilience are important paradigms for guiding decisions made under uncertainty, in particular decisions about how to protect systems from a portfolio of threats. The term *paradigm* in this context can be defined as conceptual frameworks or ways of thinking. The risk paradigm tends to emphasize reducing the probabilities and magnitudes of potential losses. The resilience paradigm tends to emphasize increasing the ability of systems to retain critical functionality by absorbing the disturbance, adapting to it, or recovering from it.

This paper addresses the applicability of these two paradigms for understanding and addressing threats that may be unknown, unquantifiable, systemic, and unlikely/catastrophic. For such threats, resilience has been favored over risk in a series of papers (Linkov et al. 2013a; 2013b; 2014a; 2014b; Park et al. 2013; Roege et al. 2014). These papers offer many insights to the study of risk and resilience and provide examples of how the study of resilience has been productive for the risk analysis community (see also Aven 2011; Baum and Handoh 2014; Haimes 2009a; 2009b; Whitten et al. 2012). However, the papers may have been too hasty to favor resilience over risk for threats that are unknown, unquantifiable, systemic, and unlikely/catastrophic. The present paper describes how both of the paradigms are comparably well suited for such threats. Instead, the paradigms are sometimes inadequately implemented for these threats, but inadequacy of implementation should not be mistaken for inadequacy of the paradigms themselves.

This paper is organized as follows. After reviewing the risk and resilience paradigms (Section 2), the paper discusses each of unknown, unquantifiable, systemic, and unlikely/catastrophic threats in turn (Sections 3-6). The discussion of these threats is illustrated using the Black Death plague in Venice, artificial intelligence (AI), and extraterrestrials. These threats are introduced in Section 3 and revisited in Sections 4-6. The Venice example suggests similar efficacy for both risk and resilience for certain unknown, unquantifiable, systemic, and unlikely/catastrophic threats. The AI and extraterrestrials examples suggest how certain unknown, unquantifiable, and unlikely/catastrophic (though less systemic) threats may sometimes be less amenable to the resilience paradigm and more amenable to the risk paradigm. Section 7 concludes the paper.

2. *Risk* and *Resilience* Paradigms

The concepts of risk and resilience have each been defined in multiple ways. A prominent definition of risk comes from Kaplan and Garrick (1981), who define risk as the triplet of possible threats, the probabilities of the threats occurring, and the magnitudes of their consequences if they do occur. The risk paradigm thus involves identifying threats, analyzing their probabilities and magnitudes, and seeking means of reducing both probabilities and magnitudes. The risk paradigm sometimes also considers potential gains in addition to potential losses, but this is less common. A prominent definition of resilience comes from the National Academy of Sciences, which defines resilience as "the ability to prepare and plan for, absorb, recover from, and more successfully adapt to adverse events" (NRC 2012:1). The resilience paradigm thus involves protecting systems from the impacts of threats so as to ensure that critical system functionality is preserved, even if it means adapting other system attributes to the changed circumstances brought by the impacts of the threats.

It has been claimed that the risk paradigm is poorly suited, and the resilience paradigm is well suited, for cases in which four conditions are met:

1) Threats are unknown. For example, Park et al. (2013:359) write that "where hazards are unknown, risk analysis is impossible", and additionally that "Resilience approaches… require preparing for the *unexpected*, whereas risk analysis proceeds from the premise that hazards are identifiable" (emphasis original).
2) Threat probabilities and magnitudes cannot readily be quantified. For example, Park et al. (2013:359) write that "even when hazards can be identified, a risk-based approach emphasizes understanding of *probabilities* of harm that may be unknowable" (emphasis original). Linkov et al. (2013a:10108) write that "resilience has a broader purview than risk and is essential when risk is incomputable".
3) Threats are systemic, targeting multiple specific system components and/or with significant effects on the rest of the system or other connected systems. For example, Linkov et al. (2014a:408) write "Unlike risk-based design, which focuses on one component at a time, resilience engineering identifies critical system functionalities that are valuable to stakeholders and society".
4) Threats are unlikely/catastrophic. For example, Park et al. (2013:359) write that "in some known, low-probability, high-consequence events… the traditional risk analysis approach has been unsatisfactory". Park et al. (2013:360, Table I) further write that while risk management aims for "minimization of probability of failure, albeit with rare catastrophic consequences and long recovery times", resilience aims for "minimization of consequences of failure, albeit with more frequent failures and rapid recovery times".

Because resilience increases the ability of systems to handle disturbances in general, it often protects systems against a range of known and unknown threats. Thus when threats are not known or cannot readily be characterized or quantified, resilience is argued to be the more suitable paradigm. However, taking a closer look at each of these four conditions shows that the four reasons for favoring resilience are mistaken. The risk paradigm is up to the task *when properly implemented*, and the resilience paradigm on its own may be inadequate for guiding decisions about protecting systems.

3. Unknown Threats

When a threat is completely unknown prior to its occurrence, risk analysis is indeed impossible. An analyst cannot estimate probabilities and magnitudes of something that lies completely outside her imagination. Likewise, she cannot do anything to manage this risk. But this fact does not make resilience any more suitable for such threats. When a threat is completely unknown, the resilience paradigm is as useless as the risk paradigm. A system manager cannot increase the resilience of a system to a threat without knowing something about how the threat would affect the system.

However, for something to be completely unknown, there must be literally zero available information about it. This is an extremely high standard. In practice, it is often possible to identify some information about threats that seem unknown. Such threats are not unknown—they seem unknown but are actually known to at least a minimal, nonzero extent. The nonzero information available about these threats makes it feasible to apply both the risk and resilience paradigms to the threats. For example, Joshi and Lambert (2011) use diversification for the management of unknown risks.

3.1 Example: The Plague in Venice

An illustration of unknown threats can be found in the case of Venice's response to the Black Death plague, which is introduced as an example of risk and resilience management by Linkov et al. (2014b). The plague arrived in Venice around 1347 and promptly caused many deaths. The Venetians at that time had no awareness of the germ theory of disease and thus initially responded in a way that can today be described as tragically mistaken:

> Those who identified the wrath of an angry God as the cause prayed and practiced self-flagellation to repent for their sins. Those who saw the sores and blisters on the skin as vulnerable openings in the body rubbed the wounds with metals commonly believed to improve protection. And those who concluded that the spread of the disease was the work of vampires buried victims with a brick wedged in the jaw to prevent the dead—now vampires themselves—from chewing their way out of the grave (Linkov et al. 2014b:378-379).

Over time, the Venetians refined their plague-fighting techniques. They isolated ships at outer islands for up to 40 days—this duration spawned the word *quarantine*. Doctors avoided contact with plague patients and developed dedicated professional communities for treating the plague. These and other measures helped reduce the severity of the plague, even though the Venetians still did not know that the plague was spread by the *Yersinia pestis* bacteria.

But *Yersinia pestis* was unknown to the Venetians only in a limited sense. The Venetians did not know that the plague was spread by a microscopic organism, since the germ theory of disease was not proposed in Europe until the 1500s. They did not know that the organism was a bacteria. They did not know that antibiotics could be developed to thwart the spread of bacteria. Or at

least there is no evidence that they knew any of these things. But they did know some important things about it. They knew that there was something out there causing a disease. They knew that there was (or at least that there might be) a mechanism through which the disease spread from person to person. And they knew that incoming ships could bring the disease. All of these facts are important pieces of information about *Yersinia pestis*. And so *Yersinia pestis* was not completely unknown.

Because *Yersinia pestis* and the disease it caused were not completely unknown to the Venetians, they were able to respond to it, and respond using elements of both the risk and resilience paradigms. Their various efforts increased Venice's resilience to the plague by enabling Venice to better prepare for incoming ships containing the plague, and by better absorbing, recovering from, and adapting to the plague-containing ships that did come in to Venice. Likewise, the Venetians did what they could to reduce the probability of specific Venetian people catching the plague by (among other things) isolating people who might already have it, and they did what they could to reduce the severity consequences of the disease by (among other things) providing medical services to plague patients. These risk management successes were achieved despite a very incomplete understanding of *Yersinia pestis*.

Linkov et al. (2014b) do not recognize that the Venetians were practicing risk management while they were *also* increasing resilience. Instead, Linkov et al. (2014b) associate the Venetian's earlier failures to the shortcomings of the risk paradigm and their later successes to the merits of the resilience paradigm. This leads to underrating the risk paradigm and possibly overrating the resilience paradigm. Without any information about a threat, neither paradigm can be applied, but when any nonzero amount of information is available, both paradigms can be applied.

3.2 More Examples: AI and Extraterrestrials

Two ongoing threats that come closer to being actually unknown are AI and extraterrestrials, or rather certain AI and extraterrestrials scenarios. For AI, the relevant scenarios are those in which a potential future "superintelligent" AI outsmarts humanity and takes over the world (Good 1965; Eden et al. 2013; Bostrom 2014). Similarly, for extraterrestrials, the relevant scenarios are those in which humanity encounters extraterrestrials that are more powerful than itself, and the extraterrestrials take over the world (Michaud 2007; Baum et al. 2011a). Both scenarios are somewhat speculative, which makes them good examples of relatively unknown threats.

For both scenarios, increasing resilience is of little use. If humanity loses control of the planet, then traditional means of increasing resilience—such as creating redundant networks, stockpiling resources, or planning to adapt and recover—do not help humanity retain its critical functionality. This holds for any reasonable definition of humanity's critical functionality: humanity's population, its civilization, and even its very existence are all threatened. The situation here is much like the situation of those many species on Earth now extinct due to their encounter with the vastly more powerful human species, or the situation of those species that would now be extinct except that humanity chose to keep them alive. For all such species, resilience does not help. So too for humanity in the face of vastly more powerful AI or extraterrestrials.

Both threats are poorly known, even if they are not completely unknown. At this time, it is not known whether it is possible to build such an AI, let alone which AI will be built and what that AI would be like. Some leading AI researchers express skepticism that such AI is possible (e.g., Horvitz and Selman 2009). Expert surveys indicate widely varying and conflicting

projections about if and when such an AI would occur, and what the consequences would be (Baum et al. 2011b; Armstrong and Sotala 2012; Müller and Bostrom forthcoming). The threat of extraterrestrials may be even less well known. It is not known whether extraterrestrials exist, or, if they do exist, whether it is possible for humanity to encounter them. It is likewise not known which extraterrestrials humanity would encounter and what those extraterrestrials would be like. All that is known is that no extraterrestrial encounter has previously occurred. Many explanations have been proposed for why no extraterrestrial encounter has previously occurred, the so-called Fermi paradox (Webb 2002). Likewise, speculations abound on what would happen if an extraterrestrial encounter occurs, though there is limited basis for assessing which of these are most likely (Michaud 2007; Baum et al. 2011a).

The examples of AI and extraterrestrials are threats that are relatively unknown, yet they may not warrant a response of increasing resilience. Instead, the only viable response is to decrease the probability of the threat manifesting. For AI, this can be done by abstaining from building potentially dangerous types of AI (Joy 2000) or by seeking to build AIs that would not harm humanity (Yudkowsky 2011). For extraterrestrials, this can be done by abstaining from transmitting messages towards parts of the galaxy likely to house extraterrestrials (Brin undated; Haqq-Misra et al. 2013) or, eventually, by abstaining from traveling around outer space. These various response options would all decrease the risk from these relatively unknown threats, even though they do not increase resilience.

4. Unquantifiable Threats

Some threats are known to exist but resist quantification. Their probabilities and/or their magnitudes are deemed unquantifiable. If the threat probabilities and magnitudes actually were unquantifiable, then calculating risk would be impossible, at least assuming that risk is calculated per the standard probability-times-magnitude formulation. Some treatments of risk do not require full quantification, for example the study of Karvetski and Lambert (2012) on risk analysis under deep uncertainty. But the threats are not entirely unquantifiable. Instead, they only seem unquantifiable. The situation here is much like the one about unknown threats. For something to be completely unquantifiable, there must be zero available information about what its quantity might be. As with the unknown, this is an extremely high standard, and one that often does not exist in practice even when it is believed to exist. The partial quantifiability makes it feasible to apply both the risk and resilience paradigms.

4.1 Example: The Plague in Venice

In the case of Venice plague risk, it might seem that the Venetians could not quantify the probability of, for example, an incoming ship transmitting new cases of the disease. But while they may not have been able to quantify this probability with any high degree of precision, they could quantify it at least to an extent. They knew the probability was greater than zero. More importantly, they knew that it was significantly greater than zero, e.g. greater than one in a million, or maybe even one in a thousand or one in some-lower-number. Likewise, they did not know the severity of the health impact that would ensure if a ship did transmit new cases of the disease. They did not even have modern public health concepts like disability adjusted life years. But they did know that that the health impact could be greater than zero. Indeed, they knew it could be the death of a nontrivial portion of Venetians. Putting these crude probabilities and magnitudes together was plenty for the Venetians to make some sound risk management decisions.

Meanwhile, sound decision making about resilience is every bit as dependent on probabilities and magnitudes. This is because efforts to increase resilience can be costly, just like efforts to reduce risk, and furthermore because these efforts are not necessarily worth the cost. Society should not make expensive investments to increase resilience against threats that are too unlikely to occur or against threats whose consequences would be too mild to justify the expense. For example, the Venetians should not have increased their resilience to the plague by blockading their island chain to prevent all incoming ships from approaching. A blockade would have made the Venetians virtually impervious to the plague, but in the process would have destroyed their livelihood. Alternatively, their livelihood—trade—could be interpreted as the critical functionality for resilience management to protect. In this interpretation, a complete blockade would be inappropriate. Instead, resilience management could favor a partial blockade, or other severe restrictions, leaving only the minimum trade necessary to maintain critical functionality. Either way, the probabilities and magnitudes of the plague were not large enough to justify such a drastic measure.

Thus, it is not necessarily the case that "In the face of these unknowns [regarding the quantities of probabilities and magnitudes], building resilience becomes the optimal course of action" (Linkov et al. 2014a:407). Resilience analysis on its own is insufficient for sound decision making. To succeed, resilience analysis needs probabilities and magnitudes that are in some way quantifiable. This is not to impose a risk perspective on the resilience paradigm. Probabilities and magnitudes are normative primitives of inherent decision making relevance. Resilience analysis would need to be supplemented with probabilities and magnitudes even if the risk paradigm had never been invented. The importance of probabilities and magnitudes is seen, for example, in the case of Venice increasing its resilience through blockade.

To be sure, risk analysis on its own is also insufficient for sound decision making. The risk paradigm typically focuses narrowly on potential losses, neglecting potential gains. Decision making should consider both. Some would argue for other factors besides gains and losses, such as categorical rules, to factor into decision making. But by at least factoring in quantified probabilities and magnitudes of threats, the risk paradigm comes closer to providing sufficient information for sound decision making.

4.2 More Examples: AI and Extraterrestrials

It may likewise seem difficult to quantify the probabilities of AI and extraterrestrials scenarios. The reasons for this are the same reasons why these threats are relatively poorly known (Section 3.2). One can scarcely quantify the probabilities of scenarios that one barely has any information about. Probabilities of AI scenarios could plausibly be derived from expert survey data, but such probability estimates are likely to be incorrect, due to the wide divergence in expert opinion and the poor track record of historical AI expert predictions (Crevier 1993; Baum et al. 2011b; Armstrong and Sotala 2012; Müller and Bostrom forthcoming). Probabilities of ET encounter scenarios could plausibly be derived from the Drake equation, which is used to estimate the number of intelligent civilizations in the galaxy, but equation parameters are highly uncertain yielding estimates that span many orders of magnitude (Wallenhorst 1981; Ćirković 2004). Finally, the fact that both scenarios are somewhat speculative can be accounted for by factoring in the probabilities that the theories supporting the scenarios are correct (Ćirković 2012).

At the same time, these probabilities are not completely unquantifiable. First, the probabilities are not zero or one. More importantly, it is reasonable that, in the next 100 years, the AI scenarios are more likely to occur than the extraterrestrials scenarios. This is because AI

technology is rapidly progressing, whereas extraterrestrials are likely much older. In other words, it is reasonable that AI scenarios would happen in this particular century, but not previous centuries, whereas it is less reasonable that extraterrestrials would happen to appear this century. Even if extraterrestrials are alerted to human civilization due to human-caused radio transmissions or other activities, the extraterrestrials likely would not appear until much later, due to the great distances they would have to travel from other corners of the galaxy. Regardless of the details, the point is that both threats are not completely unquantifiable, even if they are relatively difficult to quantify. And as discussed in Section 3.2, both threats are more suited to the risk paradigm than to the resilience paradigm.

5. Systemic Threats

Some threats threaten multiple system components or even multiple systems. When risk analysis and risk management only consider one component at a time, they are bound to perform poorly. When attention to resilience prompts analysts and managers to treat threats more systemically, this will often yield better results.

However, it is important to distinguish between the risk and resilience paradigms as they are sometimes practiced and the paradigms as they exist in theory. In theory, both paradigms can handle systemic threats. In some practice, they do. In other practice, they do not. In particular, some risk practice focuses narrowly on components when it should be more systemic. Linkov et al. (2014b:379) identify this problem in an observation that "risk assessment has been primarily focusing on the physical domain of the system, while the information, cognitive, and social domains are often ignored". The solution, however, is not to shift from risk to resilience, but to practice risk more systemically—for example, by risk assessment paying attention to the information, cognitive, and social domains.

Risk analysis practice is indeed often not systemic. The problem can be seen, for example, in risk analysis of global catastrophes (Baum et al. 2013). The risk paradigm often leads analysts to think in reductionist, non-systemic terms. This tendency of risk analysis is unfortunate. To the extent that resilience prompts more systemic thinking, analysts should in many cases use the resilience paradigm. That said, systemic risk analysis and risk management is quite feasible, even if it is not always practiced. In this direction, Haimes (2009a; 2009b) develops and advocates a systems approach to risk.

5.1 Example: The Plague in Venice

The Venetians had some understanding of systemic risk. They addressed the plague not just by treating individual patients, but also by treating the shipping systems through which new patients were produced. Modern-day analogs abound, including for new disease outbreaks, even if risk analysts do not always act accordingly.

Meanwhile, it is feasible for resilience analysis and management to focus narrowly on specific system components. The Venetians could have responded to the plague, for example, by providing medical treatment to just one infected person. Such an action may have made that one person more resilient to the plague, but it would have done negligibly little to increase Venice's systemic resilience to the plague. So both risk and resilience can be either narrow or systemic. When threats are systemic, the important thing is to analyze and manage them as such, regardless of whether the risk or resilience paradigm is being used.

Indeed, while the solutions in Linkov et al. (2014b) are expressed in terms of the resilience paradigm, the same solutions also make for good risk management. For example, their

observations of Venice handling of the plague can readily be interpreted in risk terms, as discussed above. More generally, Linkov et al. (2014a:409) argue that "early integration of resilience into the design of systems and the regulatory structures of systems management is needed to address the emerging issues associated with complexity and uncertainty". This direction could greatly help manage and reduce a variety of risks.

5.2 More Examples: AI and Extraterrestrials

The threats from AI and extraterrestrials are not appropriate examples here, since they are not particularly systemic. If an AI or extraterrestrial civilization would take over the world, the specifics of the world system are unimportant, because humans would have no ability to manage the system. Thus the appropriate response to the AI and extraterrestrials threats is not to increase the resilience of affected systems but to reduce the probability of the systems being affected in the first place.

6. Unlikely/Catastrophic Threats

Some threats are unlikely to occur, but if they do occur, the consequences would be catastrophic. The issue here is similar to that for systemic threats. When risk analysis and risk management neglect these threats, they are bound to perform poorly. When attention to resilience prevents these threats from being neglected, better results will often accrue.

The issue here is likewise similar to that for systemic threats, rooted in the distinction between theory and practice. There is nothing inherent to the risk paradigm that requires neglecting unlikely/catastrophic threats. To the contrary, there is a significant literature using the risk paradigm for the analysis and management of such threats, often using the term *extreme events* (e.g., Bier et al. 1999; Tsang et al. 2002; Zhou et al. 2012), and there is a significant literature using the risk paradigm to argue that these are often the most important threats to address (e.g., Matheny 2007; Posner 2004). The reasoning is straightforward: If risk is calculated as probability times consequence, then low probability risks can be very important if the probability is sufficiently high. Park et al. (2013:359) are correct in stating that "systematic bias in risk analysis… *can* lead to underestimation or even ignorance of such risks" (emphasis added). But when risk analysis neglects these risks, it is an error of practice, not an error of theory.

It is true that risk management practice often neglects unlikely threats even if they are catastrophic. This occurs in the widespread use of *de minimis* thresholds in risk regulation (Adler 2007). Even when *de minimis* thresholds are not specified, the risk of unlikely events is often underestimated due to psychological biases (Weber 2006). A similar situation occurs in the dismissal of scientific theories that are perceived as unlikely but, if true, have catastrophic implications (Ćirković 2012). However, when risk management practice neglects unlikely/catastrophic threats, it can be corrected through better risk management practice without reference to resilience.

Meanwhile, practice of the resilience paradigm can also be accused of neglecting unlikely/catastrophic threats. Indeed, resilience research and practice has traditionally focused on local-scale threats. The highest consequence threats are the global catastrophes, which include natural threats, such as supervolcano eruptions, and human-made threats, such as nuclear war; AI and extraterrestrials can also be counted among these threats. The global catastrophes are only just beginning to be studied in resilience terms, and by researchers motivated by the risk paradigm (e.g. Maher and Baum 2013; Baum and Handoh 2014). Of note is an analysis by Jebari

(2014) of unknown global catastrophic risks (or existential risks, in terminology of that paper). While this analysis is not framed in terms of resilience, it is in a similar spirit.

6.1 Example: The Plague in Venice

The Black Death plague was undoubtedly catastrophic. However, the story told in Linkov et al. (2014b) shows a high probability threat, because the story depicts the Venetians' reactions *after* the plague arrived. By that point, the probability of the plague arriving was one. Furthermore, the probability of additional infections was also known to be high, based on the high infection rates already observed. This part of the Venice/plague story is thus not a good example of an unlikely/catastrophic threat. (Linkov et al. 2014b do not describe it as unlikely/catastrophic.)

The plague threat *before* it arrived may have been viewed as unlikely/catastrophic by the Venetians. The Venetians were clearly caught off guard by the plague, as evidenced by their tragically mistaken initial response (Section 3.1). One can speculate on why they were caught off guard. Perhaps they never considered the possibility of a severe disease outbreak. Perhaps they considered the possibility, but dismissed it as impossible, or as too unlikely to merit attention. Or perhaps they believed it merited attention, but lacked the tools to prepare, instead resolving to react if such an outbreak were to occur. All of these possible responses mirror responses found today to contemporary unlikely/catastrophic threats, though today the tools to prepare are far more advanced.

If the Venetians dismissed the plague as too unlikely to merit attention, this would be consistent with the Park et al. (2013:360, Table I) claim that risk management aims for "minimization of probability of failure, albeit with rare catastrophic consequences and long recovery times". Instead of worrying about unlikely catastrophes, the Venetians could have better minimized the probability of failure by focusing on higher probability, less harmful events. But it would have been bad risk management, a failure to weight possible threats by their severity. Thus the issue here would be the implementation of the risk paradigm, not the paradigm itself.

6.2 More Examples: AI and Extraterrestrials

The AI and extraterrestrials threats are undoubtedly catastrophic. However, it is difficult to say whether they are unlikely, because their probabilities are so difficult to quantify. Here, the risk paradigm can be quite useful, in particular the quantification of risk as probability times severity of impacts. For both threats, the impacts could be so catastrophic that they merit attention even if their probability is low. Exactly how much attention could depend on the probability, which ties back into questions of quantifiability (Section 4). But regardless of the specifics of any quantifications, it is evident that the risk paradigm would direct at least some attention to these possibly-unlikely/definitely-catastrophic threats, whereas the resilience paradigm has little to offer.

7. Conclusions

The risk paradigm has been criticized for performing poorly when threats are unknown, unquantifiable, systemic, and/or unlikely/catastrophic. But threats are rarely completely unknown or unquantifiable. And while use of the risk paradigm may indeed sometimes perform poorly for systemic or unlikely/catastrophic threats, this is due to inadequate use of the risk paradigm, not because the risk paradigm itself is inadequate. To the contrary, the risk paradigm is quite capable of analyzing and managing these threats.

The resilience paradigm has likewise been praised for performing better for threats that meet these four conditions. Indeed, use of the resilience paradigm often does perform well for such threats. But this is due to how the resilience paradigm is used, not to the paradigm itself. The resilience paradigm may not be any better than the risk paradigm at analyzing and managing these threats. Furthermore, resilience analysis is an insufficient basis for decision making. It must be supplemented with information about the probabilities and magnitudes of threats. Risk analysis should likewise be supplemented with information about possible gains, and any other decision-relevant information, but in quantifying probabilities and magnitudes, risk analysis comes closer to providing sufficient information for sound decision making.

Ultimately, what is most important is whether threats are successfully addressed, and systems are protected, not which paradigms are used. The risk and resilience paradigms should be treated as tools to be used in service of addressing threats. The risk and resilience communities both have plenty to learn from each other. In the interest of addressing the threats, the literature discussion of risk and resilience is important to continue.

Acknowledgments

I thank Tony Barrett, James H. Lambert, and three anonymous reviewers for helpful comments on earlier versions of this paper. Any remaining errors or other shortcomings are those of the author.

References

Adler M (2007) Why De Minimis? University of Pennsylvania, Institute for Law & Economics, Research Paper No. 07-12

Armstrong S, Sotala K (2012) How we're predicting AI–or failing to. In: Ircing P, Zackova E, Polak M, Schuster R (eds) Beyond AI: Artificial Dreams. Pilsen: University of West Bohemia, pp 52–75

Aven T (2011) On some recent definitions and analysis frameworks for risk, vulnerability, and resilience. Risk Analysis 31(4):515–522

Baum SD, Haqq-Misra JD, Domagal-Goldman SD, 2011a. Would contact with extraterrestrials benefit or harm humanity? A scenario analysis. Acta Astronautica 68(11-12):2114–2129

Baum SD, Goertzel B, Goertzel TG (2011b) How long until human-level AI? Results from an expert assessment. Technological Forecasting and Social Change 78(1):185–195

Baum SD, Maher TM Jr, Haqq-Misra J (2013). Double catastrophe: Intermittent stratospheric geoengineering induced by societal collapse. Environment, Systems and Decisions 33(1):168–180

Baum SD, Handoh IC (2014) Integrating the planetary boundaries and global catastrophic risk paradigms. Ecological Economics 107:13–21

Bier VM, Haimes YY, Lambert JH, Matalas NC, Zimmerman R (1999). A survey of approaches for assessing and managing the risk of extremes. Risk Analysis 19(1):83–94

Bostrom N (2014) Superintelligence: Paths, dangers, strategies. Oxford: Oxford University Press

Brin D (undated) Shouting at the cosmos: How SETI has taken a worrisome turn into dangerous territory. http://www.davidbrin.com/shouldsetitransmit.html

Ćirković MM (2004) The temporal aspect of the Drake equation and SETI. Astrobiology 4(2):225–231

Ćirković MM (2012) Small theories and large risks—Is risk analysis relevant for epistemology? Risk Analysis 32(11):1994–2004

Crevier D (1993) AI: The Tumultuous History of the Search for Artificial Intelligence. New York: Basic Books
Eden AH, Moor JH, Soraker JH, Steinhart E (2013) Singularity Hypotheses: A Scientific and Philosophical Assessment. Berlin: Springer
Good IJ (1965) Speculations concerning the first ultraintelligent machine. In: Alt FL, Rubinoff M (eds) Advances in Computers. Academic Press, pp 31–88
Haimes YY (2009a) On the definition of resilience in systems. Risk Analysis 29(4):498–501
Haimes YY (2009b) On the complex definition of risk: A systems-based approach. Risk analysis 29(12):1647–1654
Haqq-Misra J, Busch MW, Som SM, Baum SD (2013) The benefits and harm of transmitting into space. Space Policy 29(1):40–48
Horvitz E, Selman B (2009) Interim Report from the Panel Chairs, AAAI Presidential Panel on Long-Term AI Futures. http://www.aaai.org/Organization/Panel/panel-note.pdf
Jebari K (2014) Existential risks: Exploring a robust risk reduction. Science & Engineering Ethics, forthcoming, DOI:10.1007/s11948-014-9559-3
Joshi NN, Lambert JH (2011) Diversification of engineering infrastructure investments for emergent and unknown non-systematic risks. Journal of Risk Research 14(4):1466–4461
Joy B (2000) Why the future doesn't need us. Wired 8(04):238–262
Kaplan S, Garrick BJ (1981) On the quantitative definition of risk. Risk analysis 1(1) 11–27
Karvetski CW, Lambert JH (2012) Evaluating deep uncertainties in strategic priority-setting with an application to facility energy investments. Systems Engineering 15(4):483–493
Linkov I, Eisenberg DA, Bates ME, Chang D, Convertino M, Allen JH, Flynn SE, Seager TP (2013a) Measurable resilience for actionable policy. Environmental Science & Technology 47(18):10108–10110
Linkov I, Eisenberg DA, Plourde K, Seager TP, Allen J, Kott A (2013b) Resilience metrics for cyber systems. Environment Systems and Decisions 33(4):471–476
Linkov I, Bridges T, Creutzig F, Decker J, Fox-Lent C, Kröger W, et al. (2014a) Changing the resilience paradigm. Nature Climate Change 4(6):407–409
Linkov I, Fox-Lent C, Keisler J, Della Sala S, Sieweke J (2014b) Risk and resilience lessons from Venice. Environment Systems and Decisions 34:378–382
Maher TM Jr, Baum SD (2013) Adaptation to and recovery from global catastrophe. Sustainability 5(4):1461–1479
Matheny JG (2007) Reducing the risk of human extinction. Risk Analysis 27(5):1335–1344
Michaud MAG (2007) Contact with Alien Civilizations: Our Hopes and Fears about Encountering Extraterrestrials. New York: Copernicus Books
Müller VC, Bostrom N (forthcoming) Future progress in artificial intelligence: A poll among experts. In: Müller VC (ed) Fundamental Issues of Artificial Intelligence. Berlin: Springer
NRC (National Research Council) (2012) Disaster Resilience: A National Imperative. Washington, DC: The National Academies Press
Park J, Seager TP, Rao PSC, Convertino M, Linkov I (2013) Integrating risk and resilience approaches to catastrophe management in engineering systems. Risk Analysis 33(3):356–367
Posner R (2004) Catastrophe: Risk and Response. Oxford: Oxford University Press
Roege PE, Collier ZA, Mancillas J, McDonagh JA, Linkov I (2014). Metrics for energy resilience. Energy Policy 72:249–256
Tsang JL, Lambert JH, Patev RC (2002) Extreme event scenarios for planning of infrastructure projects. Journal of Infrastructure Systems 8(2):42–48

Wallenhorst SG (1981) The Drake equation reexamined. Quarterly Journal of the Royal Astronomical Society 22:380–387
Webb S (2002) If the Universe is Teeming with Aliens—Where is Everybody? Fifty Solutions to the Fermi Paradox and the Problem of Extraterrestrial Life. New York: Springer-Verlag
Weber EU (2006) Experience-based and description-based perceptions of long-term risk: Why global warming does not scare us (yet). Climatic Change 77(1-2):103–120
Whitten SM, Hertzler G, Strunz S (2012) How real options and ecological resilience thinking can assist in environmental risk management. Journal of Risk Research 15(3):331–346
Yudkowsky E (2011) Complex value systems in Friendly AI. In Schmidhuber J, Thórisson KR, Looks M (eds) Artificial General Intelligence: 4th International Conference Proceedings. Berlin: Springer, pp 388–393.
Zhou Q, Lambert JH, Karvetski CW, Keisler JM, Linkov I (2012) Flood protection diversification to reduce probabilities of extreme losses. Risk Analysis 32(11):1873–1887