Supporting uncertain policy decisions for global catastrophic risks

by

Scott Janzwood

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Global Governance

Waterloo, Ontario, Canada, 2021

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner: Dr. James Wilsdon, Digital Science Professor of Research Policy, University of Sheffield.

Supervisor: Dr. Thomas Homer-Dixon, Professor, Faculty of Environment, University of Waterloo.

Internal Member: Dr. Vanessa Schweizer, Associate Professor, Department of Knowledge, University of Waterloo.

Internal-external Member: Dr. Jason Thistlethwaite, Associate Professor, School of Environment, Enterprise and Development, University of Waterloo.

Other Member: Dr. Anders Sandberg, Senior Research Fellow, Future of Humanity Institute, University of Oxford.

Author’s Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The three articles in this dissertation explore the contested, multi-dimensional concept of uncertainty and how experts and decision makers collectively grapple with it at governance organizations tasked with addressing global catastrophic risks (GCRs). This project examines the foundational concept of uncertainty and then explores "decision support" dynamics at the National Aeronautics and Space Administration (NASA) and the Intergovernmental Panel on Climate Change (IPCC) – the primary knowledge brokers in the governance regimes addressing planetary defense and climate change respectively.

Article #1 begins by examining the contested, multidimensional concept of uncertainty itself. The paper presents a critical analysis of the conceptual literature on uncertainty that has become increasingly standardized behind the tripartite distinction between uncertainty *location*, uncertainty *level*, and the *nature* of uncertainty. I argue that the epistemological foundation on which this framework is built is both vague and inconsistent. Perhaps most surprising is its exclusion of the term "confidence" – which has become the dominant perspective for characterizing and communicating uncertainty in many disciplines and policy contexts today. This article reinterprets the tripartite framework from a Bayesian epistemological perspective, which views uncertainty as a mental phenomenon arising from "confidence deficits" as opposed to the ill-defined notion of "knowledge deficits" that dominates the literature. I propose a more consistent set of rules for determining when uncertainty may or may not be quantified, a clarification of the terms "ignorance" and "recognized ignorance," and an expansion of the "level" dimension to include levels of uncertainty reducibility. Lastly, I challenge the usefulness of the conventional distinction made between aleatory and epistemic uncertainty and propose a more useful distinction based on developments in the field of complexity science that highlights the unique properties of complex reflexive (i.e. human) systems.

Article #2 explores the decision support process of uncertainty reduction. "Mission-oriented" public research organizations like NASA invest in R&D to improve decision-making around complex policy problems, thus producing "public value." However, the estimation of benefits produced by such R&D projects is notoriously difficult to predict and measure – a challenge that is magnified for GCRs. This article explores how public research organizations systematically reduce key uncertainties associated with GCRs. Building off of recent literature highlighting the organizational and political factors that influence R&D priority-setting at public research organizations, this article

develops an analytical framework for explaining R&D priority-setting outcomes that integrates the key stages of decision analysis with organizational and political dynamics identified in the literature. This framework is then illustrated with a case study of the NASA planetary defense mission, which addresses the GCR of near-Earth object (asteroid and comet) impacts. The case study reveals how organizational and political factors interact with every stage in the R&D priority-setting process – from initial problem definition to project selection. Lastly, the article discusses the extent to which the case study can inform R&D priority-setting at other mission-oriented organizations, particularly those addressing GCRs.

Article #3 investigates the decision support process of uncertainty communication. The uncertainty language framework used by the IPCC is designed to encourage the consistent characterization and communication of uncertainty between chapters, working groups, and reports. However, the framework has not been updated since 2010, despite criticism that it was applied inconsistently in the Fifth Assessment Report (AR5) and that the distinctions between the framework's three language scales remain unclear. This article presents a mixed methods analysis of the application – and underlying interpretation – of the uncertainty language framework by IPCC authors in the three special reports published since AR5. First, I present an analysis of uncertainty language term usage in three recent special reports: Global Warming of 1.5°C (SR15), Climate Change and Land (SRCCL), and The Ocean and Cryosphere in a Changing Climate (SROCC). The language usage analysis highlights how many of the trends identified in previous reports – like the significant increase in the use of confidence terms – have carried forward into recent assessments. These observed trends, along with ongoing debates in the literature on how to interpret the framework's three language scales inform an analysis of IPCC author experiences interpreting and implementing the framework. This discussion is informed by interviews with lead authors from the SRCCL and SROCC. Lastly, I propose several recommendations for clarifying the IPCC uncertainty language framework to address persistent sources of confusion highlighted by the authors.

Acknowledgements

I would like to thank the people that were instrumental in making this project happen. The individuals that participated in my studies, my committee members, and all of my incredible colleagues, friends, and family members have made this process fulfilling and enjoyable from start to finish.

First, I thank the individuals that participated in my studies for generously contributing their honesty, insights, and time. This project would not have been possible without them.

I thank my supervisor, Dr. Thomas Homer-Dixon, who consistently encouraged me to follow my curiosities and helped me see what I am capable of. I would also like to thank my other two committee members, Dr. Vanessa Schweizer and Dr. Anders Sandberg, for keeping the train on the tracks.

I also thank all of the geniuses at the Future of Humanity Institute at the University of Oxford for accepting me with open arms and many thought-provoking cups of tea during my time as a visiting researcher there in the spring of 2018. Again, I extend my thanks to Dr. Anders Sandberg for making me feel at home.

My PhD research was made possible by a SSHRC Joseph-Armand Bombardier Canada Graduate Scholarship, as well as the financial support from the Balsillie School's field research fund and the University of Waterloo's Briarhurst Travel and Research Award. I acknowledge that very few PhD students in Canada receive this level of support and enjoy the security and freedom of dedicating four years to a PhD without financial stress, including many who are far more deserving than I. I am indebted to the staff at the Balsillie School of International Affairs and the University of Waterloo for their administrative support, particularly Dr. Andrew Thompson.

And I must thank my lovely friends and colleagues at the Balsillie School who made each day at the office interesting and joyful. I could not have asked for a more mutually supportive and intellectually enriching environment. In particular, I thank my fellow members of the Complexity Working Group, my procrastination accomplices and officemates, Clay Dasilva and Allison Petrozziello, my brilliant collaborator Jinelle Piereder, and my "hallway homies," Caleb Lauer, Diana Thomaz, Justine Salam, Tahnee Prior, and Johanna Wilkes.

And most of all, I want to thank my wife Amy. This dissertation, like everything in my life, is for her.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction: Supporting uncertain policy decisions for global catastrophic risks

Catastrophe seems omnipresent these days. As of January 22, 2021, the COVID-19 pandemic has killed over 2 million people. The 2017 Atlantic hurricane season was the costliest on record (NOAA 2020) and wildfires in Australia in late-2019 and early-2020 destroyed 7.7 million hectares of forest, while leading to the deaths of 33 humans and over one billion animals (Evershed et al. 2020).

Despite the enormous human, social, economic, and environmental cost of these crises, scientists warn about the possibility of risks whose impacts would be orders of magnitude higher. Global catastrophic risks (GCRs) are risks of events that could significantly harm or destroy human civilization on a global scale (Hempsell 2004b; Baum 2010; Barrett 2017). Neither the COVID-19 pandemic, nor the recent hurricanes and wildfires come anywhere close to the threshold of global catastrophe. Current GCRs that pose a potential threat to humanity include (even deadlier) pandemics, asteroid and comet impacts, climate change, nuclear war, bioterrorist attacks, and artificial intelligence (AI) accidents.

GCRs pose unique challenges for policy makers. Governments and other organizations addressing risks typically rely on expected cost calculations to guide their decision-making. Expected cost calculations integrate estimates of the potential costs imposed by a risk event with estimates of the likelihood of it occurring. However, GCRs tend to “break” expected cost calculations – a problem that stems from two factors: (1) the impact of a single catastrophic event exceeds anything we have experienced in recent human history and (2) the perceived likelihoods of various GCRs manifesting within the next century are either highly uncertain or believed to be quite low.[1] In other words, the potential costs imposed by a GCR are so great that they are potentially incalculable and their likelihood is difficult, if not impossible, to determine with a high degree of confidence.

Even when considered individually, each GCR could justify enormous investments in interventions to avoid or mitigate the threat. When considered as a part of a larger global risk portfolio that includes several GCRs (not to mention the existence of hundreds of less impactful but more likely and immediate threats like COVID-19), the fundamental limitations of our ability to manage global risk are exposed: the

[1] For a summary of recent estimates of GCR likelihoods, see Turchin and Denkenberger (2018).

number and magnitude of problems exceeds the capabilities and resources of governments and other institutions to deal with them. Not only do we lack the financial resources to fund interventions and other governance activities to effectively address every threat, we also lack the necessary time, organizational capacity, ingenuity, and attention (Homer-Dixon 2000).

While there is currently no consensus on how to best triage the risks in our global risk portfolio, a few GCRs are actively being governed. For instance, the National Aeronautics and Space Administration (NASA) spends hundreds of millions of dollars per year tracking near-Earth objects (NEOs) like asteroids and comets and developing technologies to prevent potential collisions with the Earth. Meanwhile, national governments around the world continue to claim they are pursuing policies consistent with limiting global temperature rise to 1.5°C above pre-industrial levels;[2] a target which is believed to significantly reduce the likelihood of catastrophic scenarios (IPCC 2018).

However, in both contexts, decision-making about what actions should be taken and when to take them is handicapped by pervasive and deep uncertainty about the nature of the risks themselves, uncertainty about possible solutions to eliminate or decrease the risk, and uncertainty about how to effectively implement those solutions. Therefore, a substantial amount of the governance activity addressing GCRs is dedicated to two tasks: reducing and communicating uncertainty. The process whereby experts take actions to improve the ratio of “what we know” to “what we do not know” and then communicate policy-relevant knowledge and knowledge gaps to decision makers is commonly referred to as *decision support*.

Decision support is the process through which knowledge producers like scientists and other experts influence the decisions of policy makers and other decision makers like politicians and business leaders. Often, this relationship is more-or-less transactional, with experts supplying policy-relevant knowledge to meet the demand of decision makers. Sometimes, experts are employed by the same organization as the decision makers (e.g. government scientists). Other decision support relationships are more consultative, with experts communicating the knowledge they produce as employees of universities or other research organizations to policy makers. Like in any market, the levels of supply and demand do not always match. Decision makers may desire answers that experts are unable or unwilling to provide, while knowledge produced by experts may be ignored.

While useful, this transactional framing of decision support fails to address the complex, bidirectional nature of many real-world interactions between experts and decision makers. These dynamics are

[2] A target that most experts believe to be difficult, if not impossible to achieve (Raftery et al. 2017).

captured by the concept of *co-production* (Jasanoff 2004), which positions decision support as a dynamic process where influence, information, and values are communicated back-and-forth across the so-called "science-policy boundary." From this perspective, policy-relevant knowledge does not emerge miraculously within the mind of an expert before being passed along to a receptive decision maker – but rather, it is co-produced by the context-specific interactions of various actors within and between complex organizations.

This dissertation explores the contested, multi-dimensional concept of uncertainty and how experts and decision makers collectively grapple with it at governance organizations tasked with addressing GCRs. The three overarching research questions I address are:

1. What is uncertainty? More specifically:
    a. What does it mean to *reduce* uncertainty?
    b. What does it mean to *communicate* (i.e. characterize, measure, describe) uncertainty?
2. How are organizations that address GCRs systematically reducing uncertainties to support decision-making?
3. How are organizations that address GCRs systematically communicating uncertainties to support decision-making?

This dissertation contains an introduction, three articles, and a brief conclusion. In the remainder of the introduction, I provide more background on model-based decision support and GCRs, I summarize the contributions made by each article, and I outline the structure of the dissertation.

1.1 Decision support

1.1.1 Theoretical perspectives on decision support

The perspective from which I explore how experts and decision makers collectively confront uncertainty is informed by the broader theoretical literature on the relationship between scientific knowledge production and political decision-making. A useful starting point is the four archetypal models of the science-policy relationship proposed in the sociological literature: the technocratic model, the decisionist model, the fatalist model, and the pragmatist model (Habermas 1971; Edenhofer and Kowarsch 2012).

The technocratic model sees policy problems as entirely scientific in nature and therefore, scientists should define the problem, identify possible solutions, and provide an assessment of the best solution, including the specific policies and technologies necessary to implement it. Meanwhile, the decisionist model also sees the framing of the problem and the identification of possible solutions as largely objective and value-neutral but acknowledges that evaluations of solutions, policies, and technologies are unavoidably value-laden and thus lie beyond the scope of science. Rather, it is the responsibility of policy makers to make these more subjective and normative decisions.

The fatalist model takes the opposite position of the technocratic model, viewing science as one of many equally valid types of knowledge. From this perspective, scientific knowledge possesses no inherent authority or legitimacy, and its status in the eyes of policy makers is entirely the product of power relations. This perspective is a useful (although overly simplistic) lens from which to view the rise of anti-expert sentiment in recent populist authoritarian political movements around the world. The pragmatist model, which occupies the middle ground between the decisionist and fatalist interpretations, has the most explanatory power for describing the complex, messy decision support relationships found in most policy contexts today. According to this model, all policy-relevant knowledge is imbued with values and cannot be completely separated from the social and cultural context from which it is produced. However, scientific knowledge still possesses a "pragmatist objectivity" that is intersubjectively negotiated between stakeholders (Edenhofer and Kowarsch 2012).

Most theoretical perspectives on the intersection between science and policy resemble the pragmatist model, where science is at once contested and negotiated but is still granted the authority of privileged knowledge. Some perspectives, such as the literature on epistemic communities, highlight the control of knowledge and information as an important dimension of power. For example, Haas examines the role of experts in "articulating the cause-and-effect relationships of complex problems, helping states identify their interests, framing the issues for collective debate, proposing specific policies, and identifying salient points for negotiation" (1992, pp. 2). Marxist and neo-Gramscian perspectives view scientific knowledge as a hegemonic tool (Aronowitz 1988; Porter 1995) and highlight the role of structural power — that is, the ability of scientists and experts to influence the capacities of political actors by directly shaping the social rules of constitution (Barnett and Duvall 2005).

Meanwhile, social constructivism also provides a useful lens for understanding the dynamic processes through which knowledge is constructed and how it shapes political outcomes. Constructivist scholars describe how the interests of experts and policy makers are derived from the existing social structure of norms and institutions in which they are enmeshed (Finnemore 1996; Katzenstein 1996; Ruggie 1998). Bernstein's (2001) socio-evolutionary approach explains the "uptake" dynamics of scientific knowledge by decision makers. He argues that the acceptance of new norms or ideas — such as the perceived validity and utility of scientific knowledge — depends on their "social fitness" within the existing institutional, social, and cultural context. This perspective sheds light on why the same piece of expert knowledge may differently impact political decision-making in one organizational context than it does in another.

These theoretical perspectives combine to paint a picture of the science-policy interface that is at once dynamic, value-laden, context-dependent, and constantly negotiated by actors yielding different types of power. This perspective informs my decision to look closely at specific organizational contexts to better understand how decision support relationships function and evolve.

1.1.2 Models all the way down

All decisions – political or otherwise – require the implicit or explicit construction of models. Models are mental renderings of real-world systems that are constructed to lay bare our understanding of system dynamics and to make predictions about their behaviour. Prediction does not refer only to forecasts of the future but also, in the more general scientific sense, to the deductive logical consequence of a theory (Beinhocker 2013). Therefore, a prediction may refer to a claim about the past, present, or future.

Models define the key components and relationships that are implicated in the chain connecting causal factors to the uncertain outcomes we want to predict — like how drastically the climate will change in the next 50 years or the virulence of a recently detected infectious disease outbreak. Descriptive models can help manage uncertainty by revealing particularly uncertain aspects of the system, while predictive models can reduce or resolve uncertainty by simulating the behaviour of the real-world system.

The philosophical literature on uncertainty primarily focuses on mathematical models, which are used to describe systems whose key physical and non-physical elements can be represented numerically as parameters or variables, and their interactions and causal relationships can be defined mathematically. The systematic nature of mathematical modeling makes it particularly useful for illustrating the various dimensions of uncertainty outlined in Article #1. However, just as relevant to decision support relationships are the many less sophisticated models we use to understand systems and make decisions on a moment-by-moment basis, such as the mental models we construct to inform simple decisions like whether or not to trust someone or if we should grab an umbrella before leaving the house.

Modelers face a series of trade-offs between the level of detail they include in their model and its tractability (the cost, effort, or computational resources required to "solve" or compute the model).

Modelers must make choices about which aspects of reality to include and which to exclude in order to produce the most accurate and useful model possible. Thus, all models aim to be "useful simplifications" of the real-world system.

One consequence of this simplification is that models reflect a specific scale or vantage point. It often takes multiple models describing different aspects of the system to fully capture the system components, relationships, and behaviour that are of interest to the analyst. Take, for example, mathematical epidemic models, which are used to understand and predict the dynamics of disease outbreaks. These models identify the relevant components of the system, such as the lethality of the pathogen, the size of the exposed population, the effectiveness of interventions, and the rate of spread. However, any one of these model components can also be described by its own sub-model. For example, the rate at which an outbreak spreads (also known as the basic reproduction number) involves its own set of components and relationships that can be modeled, including: the number of confirmed and suspected cases, population mixing dynamics, pathogen phylodynamics, and the rate of susceptible-host transmission. And in turn, each of these components or relationships can also be described by its own model.

Therefore, models tend to be nested within one another like Russian Matryoshka dolls. Zooming into a disease outbreak model, we see more fine-grained models describing the biological properties of the pathogenic organism, the immune system of an individual human host, and the behaviour of particular organs and cells. Zooming out, we see models of the social and cultural systems that influence preferential mixing patterns, the technological systems that improve interventions, and the political and economic systems that influence the supply and effectiveness of those interventions. The point is that all decisions are supported by mental renderings of reality, many of which we formalize mathematically into complex computational models that become progressively more fine-grained the closer we look.

A final important point to make about model-based decision support is that while models are built to manage or resolve uncertainties, they also inject entirely new uncertainties into the decision context – uncertainties that arise from the modeling process itself. In Article #1, I discuss how scholars have divided up the modeling process into six overlapping tasks: (1) bounding the system; (2) constructing the conceptual model; (3) constructing the computer model; (4) inputting data; (5) implementing the model; and (6) communicating model outputs (Kwakkel, Walker, and Marchau 2010). In any given modeling study, key uncertainties may arise from any of these tasks. Therefore, novel uncertainties may emerge from the very strategy used to reduce or measure uncertainty in the first place, which can be equally debilitating to decision-making.

1.2 Global catastrophic risk

GCR describes the class of risks facing humanity whose potential impacts are of an extreme or unprecedented magnitude, such as a substantial decrease in global population, a significant and irreversible decrease in human welfare, the collapse of civilization as we know it, or human extinction. Experts propose that risks such as pandemics, bioterrorism attacks, NEO impacts, supervolcanoes, climate change, and AI accidents could all plausibly produce disastrous consequences for the long-term prospects of human civilization. A comprehensive list of (currently known) GCRs is provided in Table 1.

Table 1: Global catastrophic risks (sources: Global Challenges Foundation 2018; Bostrom and Ćirković 2008)

Risks from nature	*Risks from accidents*	*Risks from malicious actors*
Asteroid and comet impacts	Anthropogenic climate change	Bioterrorism
Supernovae and gamma ray bursts	Non-aligned superintelligent AI	Weaponized AI
Supervolcanoes	Nanotechnology accidents	Nuclear war/terrorism
Natural pandemics	Physics experiment accidents	
Climate variation	Biotechnology accidents	

1.2.1 Why GCRs?

Underpinning the GCR concept is an implicit claim that these risks require either special consideration or special analysis. In terms of deserving special consideration, a common refrain in the GCR literature is that GCRs are understudied within academia, insufficiently addressed by governance institutions, and underappreciated by the public at large. For example, GCR has been described as “a critical topic that has long remained understudied” (Liu, Lauta, and Maas 2018) and it has been argued that GCR researchers have a special obligation to raise awareness and strengthen public concern (Rees 2013). The claim that GCRs require special analysis is supported by the fact that GCRs are not adequately managed using expected cost analysis (Jablonowski 2007; Matheny 2007). So long as the likelihood of any of these risks is non-negligible, from an expected cost perspective, the prospect of infinite (or near-infinite) damages should prompt governments to direct most, if not all, of their resources towards avoiding a small chance of catastrophe.

The central debate over the GCR concept itself is where to set the floor separating GCRs from sub-catastrophic risks.[3] Proposed definitions of GCR disagree on where to set the "catastrophe threshold" – or whether there should be a precise threshold at all. Typically, the threshold is expressed in terms of the number of human lives lost or negatively affected, such as 10% of the global population (Cotton-Barratt et al. 2016; Turchin and Denkenberger 2018), significantly more than 10% of the global population (Hempsell 2004b), or 25% of the global population (Atkinson 1999). Other definitions appeal to intentionally imprecise or qualitative measures such as irreversible civilizational collapse (Baum 2010; GCRI 2018). Meanwhile, in the popular discourse, the term catastrophic seems to be used according to the principle of "you know it when you see it."

However, most conceptualizations seem to agree that the impacts of a GCR would be unprecedented, which would mean that the impacts of a GCR would exceed the estimated 35 to 60 million deaths (1.5-2.6% of global population) that occurred during World War II (Royde-Smith and Hughes 2019) or the 50 to 100 million deaths (2.5-5% of global population) believed to be caused by the 1918-1921 influenza pandemic (Johnson and Mueller 2002). Unprecedentedness seems to capture one of the reasons why policy makers appear to be largely oblivious of, or unconcerned about, GCRs. The other reason why the GCR research community has struggled to place GCRs more prominently on the policy agenda is the fact that GCRs are believed to be unlikely to occur within the next 100 years (Turchin and Denkenberger 2018).

It is worth noting that definitions of GCR are almost uniformly anthropocentric, focusing on harms inflicted on human populations. Therefore, according to these definitions, the unfolding sixth mass extinction (Barnosky et al. 2011) would not constitute a GCR as long as biodiversity loss and ecosystem collapse do not lead to massive economic costs or human deaths. However, there are a number of more expansive conceptualizations of GCR in the literature, including proposals to extend the measurement of disutility to species "we care about," (Torres 2015) or to all "morally relevant beings" (Bostrom and Ćirković 2008).

[3] Complicating matters further, the concept of existential risk (X-risk) has also been proposed as a sub-class of GCR describing the most extreme risks that could cause human extinction or trigger an irreversible civilization collapse (Bostrom 2002). From this perspective, the concept of GCR either serves as an overarching category of high-impact risk (e.g. Bostrom and Ćirković 2008) or as a vague intermediate category separating sub-catastrophic risks and X-risks.

1.2.2 History of the GCR research program

Scholarship on high-impact risks began in earnest in the mid-twentieth century[4] addressing the destructive capabilities of nuclear weapons. Much of the early thinking on threats posed by nuclear technologies was published in the Bulletin of Atomic Scientists, including research addressing the early fear that a nuclear explosion could ignite all the nitrogen in the atmosphere or hydrogen in the oceans (Bethe 1976).

New fears emerged in the 1980s that a sufficiently large nuclear war could trigger atmospheric and climatic effects, creating a “nuclear winter” that could be far more impactful than the immediate fallout (Turco et al. 1983). By integrating existing research on volcanic eruptions and climate effects, the research exploring the threat of nuclear winter foreshadowed the field’s emerging intuition that important lessons can be applied from one risk area to another.

While many scholars concerned with high-impact risks were preoccupied with the threat of nuclear annihilation, others drew attention to new technologies and physics experiments that, while less understood, could prove to be even more destructive than the nuclear threat. The great particle accelerator projects of the 1990s and early-2000s (the RHIC and LHC) prompted physicists to explore the possibility that high-energy particle collisions could unleash a number of unprecedented and poorly-understood physical processes such as strangelets, vacuum instability, or black holes (Dar, De Rujula, and Heinz 1999; Jaffe et al. 1999). Around the same time, early advances in nanotechnology elicited concerns that molecular assemblers could trigger runaway self-replication, destroying all biomass on Earth (Joy 2000).

Research has also been conducted on external threats, such as NEOs, including research calculating the likelihood of a “doomsday impact” (Chesley et al. 2002; Chapman 2004; Asher et al. 2005). Much of the astrophysical research has been propelled, in part, by breakthroughs in paleontological research linking the Cretaceous–Tertiary (K–T) extinction to a massive impact (Alvarez et al. 1980; Hildebrand et al. 1991). Paleoanthropological research has also played a significant role in quelling culturally dominant narratives around human progress and invincibility by revealing the historical collapse of ancient civilizations (Butzer and Endfield 2012), periodic mass extinctions (Raup and Sepkoski 1982), and the disappearance of the Neanderthal and other hominids (Higham et al. 2014).

High-consequence risks have also been of great interest to philosophers investigating the so-called “doomsday argument” and the related issue of observation selection effects. First proposed by Carter

[4] Although Moynihan (2020) traces the intellectual history of GCRs and existential risk all the way back to the eighteenth-century Enlightenment.

(1983) and advanced by Leslie (1992), the doomsday argument asserts that, as a species, it is more likely that we are closer to the end of the timeline of our existence than we are to the beginning. Other philosophers have raised the issue that such assessments of our species' future prospects tend to be skewed by an "anthropic bias," where the fact that we have (thus far) avoided extinction obscures our ability to accurately assess how likely or unlikely that may be (Tegmark and Bostrom 2005).

Up until the early-2000s, high-impact risks were typically treated as discrete research topics within established academic fields, such as astrophysics, epidemiology, and security studies. The publication of the edited volume *Global Catastrophic Risk* (Bostrom and Ćirković 2008) marked a turning point for the growing and increasingly organized research community addressing GCR. Since 2008, a large amount of the GCR literature has focused on solving two of the central puzzles at the heart of the research program: the "high/low problem" and the "discounting problem."

1.2.2.1 The high/low problem

Reconciling the potentially massive impacts of GCR with their relatively low (and deeply uncertain) probabilities continues to be the central puzzle at the heart of the GCR field. I refer to this challenge as the *high/low problem*. Conventional risk management methodologies typically involve expected cost calculations, which rely on accurate estimations of the expected impact of a risk event and the likelihood that it will occur within a specified time horizon. GCRs complicate expected cost calculations because, should a risk event occur, they may have infinite (or near-infinite) disutility (Weitzman 2008). Even if a risk event is extremely unlikely, an expected cost calculation would suggest that we should spend a near-infinite amount of money to eliminate the threat or reduce its impact. Complicating this calculation further, the probabilities of most GCRs are deeply uncertain and might actually be significantly higher than are widely believed (Taleb 2007).

Important early contributions to solving the high/low problem include research addressing the difficulty of estimating probabilities for rare or unprecedented events (Matheny 2007; Jablonowski 2007) and the development of new methodologies to measure the probability of catastrophe and eliminate anthropic bias (Ludwig 1999; Bostrom 2002; Bostrom 2009; Tegmark and Bostrom 2005; Munthe 2019). Over the last ten years, researchers have continued to examine sub-challenges related to the high/low problem including: estimating the likelihood of rare or unprecedented events, estimating the magnitude of negative consequences should a risk event occur, and making decisions when these two values are uncertain.

Likelihood estimation has been addressed by scholarship identifying a number of methodological problems associated with quantitative GCR models that produce probability estimates (Ord, Hillerbrand, and Sandberg 2010; Manheim 2018), as well as attempts to quantify subjective expert opinion using elicitation methods (Sandberg and Bostrom 2008; Grace et al. 2018). Meanwhile, Tonn and Stiefel (2013) have proposed integrating multiple methodologies to estimate extinction risks, such as scenarios, Bayesian networks, and expert elicitation.

Issues related to impact estimation have been addressed within the neighbouring literature on climate-economic modeling, which has debated whether or not we are able to bound disutility for catastrophic climate scenarios. A number of economists argue that deep uncertainties in the climate system prevent us from ruling out the existence of a fat-tailed probability distribution function (non-negligible probability values for catastrophic climate scenarios) and a highly uncertain damages function that determines the human and economic costs imposed by different scenarios (Weitzman 2008; Pindyck 2011). Responding to the uncertainty in probability estimates and damages functions, Barrett (2017) has proposed applying a modified form of expected value of perfect information (EVPI) analysis to GCRs to calculate the potential benefits from reducing uncertainty around GCRs.

It is worth mentioning that the high/low problem may have decreasing relevance to climate change – one of the two GCRs explored in this dissertation. While the prospect of a 10% decrease in global population as a (more or less) direct result of climate change may have seemed extremely unlikely even ten years ago (Sandberg and Bostrom 2008), a growing awareness of the presence of interconnected tipping elements in the climate system amongst experts (Steffen et al. 2018) may be leading to a reframing of climate change as more of a “high/medium problem.”

1.2.2.2 The discounting problem

Discounting, also referred to as social time preference, refers to how we determine the relative value of current and future welfare in order to maximize welfare in the aggregate. Economic theory asserts that since per capita welfare tends to increase over time and the marginal utility of welfare decreases, privileging short-term investments over long-term investments maximizes human welfare in the long run.

However, GCRs run into problems with conventional justifications for high discount rates. First, discount rates measure costs in dollars, while the costs imposed by GCRs are often measured in human lives. Specifically, GCR researchers have criticized value of statistical life (VSL) calculations that are commonly used by economists to convert deaths to financial costs (e.g. Barrett 2017). Second, discount rates may not apply to extinction risks because it would not make a great deal of sense to privilege current welfare and increase the "size of the pie" if there is no one around in the far-future to eat it.

The issue of how to value far-future human life given the existence of plausible extinction risks has been a particularly active topic in moral philosophical literature on GCR since Parfit's thought experiment comparing the disutility of civilizational collapse and extinction. In his book *Reasons and Persons* (1984), Parfit compares a risk event that eliminates 99% of the global population (event A) to a subsequent event that eliminates the final 1% (event B), suggesting that the very large, immediate disutility from event A pales in comparison to the loss of countless future generations in event B. This consequentialist perspective has come to dominate the GCR literature, reflecting the position that it is unethical to discriminate against spatially remote and temporally remote human lives (e.g. Sandberg 2017).

1.2.3 **Governance and global catastrophic risk**

Growing concern about GCRs has created a sense of urgency in the emerging literature on GCR governance, which has begun to sketch out the institutional and normative frameworks necessary to effectively avoid or mitigate these risks. Recent scholarship maps out the main technical and political challenges facing AI governance (Dafoe 2018a), proposes an inventory of "critical governance failures" (Liu, Lauta, and Maas 2018), and evaluates the relative cost-effectiveness of biosecurity strategies (Millett and Snyder-Beattie 2017). While much of the emerging GCR governance research has focused on imagining what effective GCR governance regimes *should* look like, there is a notable lack of scholarship mapping out what GCR governance *does* look like today. Descriptive research can complement existing normative scholarship by pointing to key deficits in GCR governance activities, as well as opportunities for innovation and change.

While the term governance is often used to describe formal rules like laws, policies, and regulations that are developed and implemented by governments and other organizations to address policy problems, governance also refers to the development of informal rules (i.e. norms), strategies, and practices (Ruggie 2014). Therefore, GCR governance is understood here as *the systems of authoritative rules, institutions, norms, strategies, and practices that actors use to avoid or decrease the impacts of GCRs*. Adopting this definition, I argue that the systematic reduction and communication of uncertainty qualify as practices performed by research organizations aimed at developing strategies to avoid or decrease the impacts of problems like GCRs.

For some GCRs like climate change, nuclear war, and pandemics, there is no shortage of literature describing the development of rules-based governance regimes. However, for risks like AI accidents, nanotechnology accidents, and NEO impacts, there is a noticeable absence of formal rulemaking. In fact, a cursory reading of the GCR literature may create the impression that little is being done to address most GCRs today. However, this view reflects the narrow interpretation of governance as formal rulemaking. When the scope of governance activities is expanded to include various practices and strategies associated with research and development (R&D) and decision support, such as the identification and framing of threats, uncertainty reduction, and the development of solutions, the current state of GCR governance is far more active than is conventionally believed.

1.2.3.1 Catastrophic vs. sub-catastrophic risk governance

The obvious starting point for mapping the current landscape of GCR governance is to identify the actors and institutions already addressing GCRs. However, there are no dedicated institutions focusing solely on very large asteroids or extreme climate scenarios. Rather, these risks are addressed by the same institutions, experts, and policy makers that address far more likely but less calamitous versions of these risks (i.e. sub-catastrophic asteroid impacts and climate change scenarios).

It is surprisingly difficult to draw a clear line separating GCR governance activities from activities addressing sub-catastrophic risks because many of the mitigation or adaptation strategies already proposed to address sub-catastrophic risks are the same strategies that would effectively address GCRs. For instance, both catastrophic and sub-catastrophic pandemic governance draw from the same basic portfolio of intervention options: surveillance, vaccine development, quarantine, and treatment.

The task of distinguishing between catastrophic and sub-catastrophic risk governance is easier for some threats than others. First, some GCRs like supernovae, gamma-ray bursts, nanotechnology accidents, and physics experiment accidents do not have a related sub-catastrophic risk. Should any of these threats occur, the most plausible outcome is human extinction. Other threats, like asteroid impacts and AI accidents have both high- and low-impact scenarios. Most experts agree that asteroids or comets at least 1 to 2 km in diameter pose a catastrophic threat, while smaller asteroids do not (Chapman 2004; Garshnek, Morrison, and Burkle 2000). For the risk of AI-related accidents, experts have proposed plausible scenarios where superintelligent AI systems with vastly greater cognitive capabilities than humans become unintentionally misaligned with human values, posing a catastrophic or existential threat (Bostrom 2014). However, many less impactful AI scenarios have also been proposed, such as concerns around labour displacement and the equitable distribution of AI capabilities and benefits (Frank et al. 2019).

In the case of both planetary defense (NEO impact prevention) and AI governance, the high- and low-impact versions of these risks are easy to distinguish at an ontological level. In other words, the causal models describing the initial conditions, factors, and outcomes is sufficiently dissimilar that experts often speak of them as separate things. For instance, the distinction between the risk posed by a 1 km-wide rock hurling towards Earth and a 10 m-wide rock is immediately apparent: the larger asteroid is a GCR, while the smaller asteroid is not. Likewise, the risk that certain countries or corporations may dominate the AI race, thereby exacerbating inequality, feels qualitatively different from the direct physical threat posed by AI systems to humans. However, in both cases, catastrophic and sub-catastrophic risks are largely addressed within the same institutions by the same assemblage of experts and policy makers. For instance, the NASA NEO Observation Program tracks catastrophic and sub-catastrophic asteroids alike, while nascent AI governance organizations like OpenAI and the Machine Intelligence Research Institute seek to address all manner of negative externalities emerging from AI development.

However, these clear ontological distinctions between catastrophic and sub-catastrophic risks begin to break down when it comes to risks like infectious disease outbreaks, bioterrorism, and climate change. For both infectious disease outbreaks and bioterrorism, there appears to be a specific set of characteristics of pathogen-host dynamics that make catastrophic impacts possible. These characteristics include: the ability for the pathogen to be spread through respiration, a high case fatality rate, efficient human-to-human transmissibility, an "immunologically naïve" population, the ability for the pathogen to be transmitted during incubation periods, and the lack of effective or available countermeasures (Adalja et al. 2018). Only infectious diseases that possess these characteristics could plausibly produce a catastrophic outcome. However, the water is muddied somewhat for a pathogen like influenza that is constantly evolving and has many strains – some of which may possess these properties and some of which do not.

The distinction between catastrophic and non-catastrophic climate change is even more tenuous. The terms "catastrophic climate change" and "dangerous climate change" frequently appear in both the scientific literature and the news media – albeit to describe impacts well below any of the discussed definitions of catastrophe. The term dangerous climate change has been used by the IPCC to describe all impacts above 2°C, and more recently 1.5°C (Pereira and Viola 2018), while the language of catastrophe is commonly invoked in the econometric literature to describe a wide range of damages (Kopits, Marten, and Wolverton 2014). Catastrophic climate change has been used to describe specific climate events at vastly different spatial and temporal scales like rising ocean levels (slow and global) or an extended drought (fast and local) (Tsur and Withagen 2013) – none of which imply truly catastrophic impacts on their own.

So, what level of climate change is necessary to produce impacts approaching the catastrophe threshold? As a starting point, the most extreme warming scenarios considered by climate models (typically 6°C) can be confidently labeled catastrophic. At 6°C, it is estimated that 50-80% of the world will encounter conditions where the Wet Bulb Globe Temperature (the temperature at which the human body can no longer maintain its core temperature in the shade) is exceeded for at least 10 percent of days in the hottest month of the year (King et al. 2015). At 6°C, it is plausible that catastrophic impacts may occur from high temperatures alone. If 6°C warming almost certainly qualifies as catastrophic climate change, then the catastrophe threshold must be at a less extreme warming level.

However, for lower (and more likely) warming scenarios like 3°C, accurately projecting impacts is complicated by a long list of uncertainties such as the level and timing of adaptation policies and the potential for latent tipping points or feedback effects in the climate system. A recent study estimating the rate of temperature-related deaths under different warming scenarios shows that first-order impacts (i.e. impacts stemming directly from temperature change) at 3°C come nowhere near the catastrophe threshold (Gasparrini et al. 2017), which seems to suggest that for 3°C to produce catastrophic impacts, it would need to come from second-order impacts.

However, analyses of second-order climate change impacts tend to provide either broad qualitative estimates of harm or quantitative estimates of impacts not expressed in terms of human mortality or welfare. For example, 3°C is expected to produce largescale ecosystem collapse, the loss of a significant portion of arable land and freshwater resources, and a significant decrease in agricultural production (King et al. 2015; IPCC 2014). 3°C has also been connected to the displacement of as many as 1 billion people from low-lying coastal regions (Watts et al. 2018), an increased rate of vector-borne diseases like malaria and dengue fever (IPCC 2018), and the possibility that 100-year-droughts could occur every 2 to 5 years (Naumann et al. 2018). However, the implications of these impacts for human welfare and mortality rates are not entirely clear.

Yet, several recent reports disagree and continue to point to 3°C as the threshold where catastrophic impacts become plausible. A 2018 report by the National Centre for Climate Restoration suggests that "adverse outcome[s] that would either annihilate intelligent life or permanently and drastically curtail its potential" become possible if the climate were to cross the 3°C threshold (Spratt and Dunlop 2018).

Specifically, the report cites a study linking 3°C of warming and a 0.5 m rise in sea-level with a heightened risk of nuclear war and social upheaval (Campbell et al. 2007). The case for catastrophic impacts becomes more compelling at 4°C, at which point 74% of the global population could be exposed to "deadly heat" (Xu and Ramanathan 2017) and our ability to effectively adapt becomes deeply uncertain (World Bank 2012).

However, the most compelling argument of why a 3°C scenario could produce catastrophic impacts is the existence of potential tipping elements like the ice-albedo feedback activated by the collapse of the Western Antarctic and Greenland ice sheets, the reversal of the thermohaline circulation in the North Atlantic, and slower moving processes like permafrost thawing and the decomposition of ocean methane hydrates (Lenton et al. 2008). In fact, it has been suggested that these tipping points could be activated at a global temperature change as low as 2°C (Steffen et al. 2018). The main implication of triggering these tipping points is that it may actually be impossible to stabilize the Earth's climate at 3°C (Xu and Ramanathan 2017; Pereira and Viola 2018; Steffen et al. 2018). In other words, there may be no 3°C scenario at all. Like a car hitting a patch of black ice while braking for a stop sign, by aiming for 2 to 3°C (the "stop line"), we may inadvertently keep sliding towards 4°C and beyond, even if emissions are significantly reduced.

Therefore, here I use the interval of 2 to 3°C as my fuzzy boundary separating catastrophic and sub-catastrophic climate change due to the non-negligible probability that feedback effects propel the climate system into extreme warming scenarios. However, if 2 to 3°C is the point when catastrophic impacts become plausible, then what exactly is catastrophic climate change *governance*? Since any activity that plays a role in preventing a global temperature increase of 1.5°C would also help prevent a 3°C scenario, one might be inclined to argue that *all* climate change governance qualifies as catastrophic climate change governance.

However, the disappointing commitments made by national governments under the Paris Agreement (which, if achieved, would stabilize the climate at around 2.7°C (Climate Action Tracker 2015)) reveal the persistence of a linear understanding of the climate system that either misinterprets or ignores the latest climate science on tipping elements. From this linear perspective, 2°C is worse than 1.5°C – and 2.5°C is worse than 2°C (by somewhat equal increments of disutility). However, from the nonlinear perspective advocated by experts warning of latent tipping elements in the climate system, 2°C is worse than 1.5°C – but 2.5°C is *significantly* worse than 2°C. Therefore, I distinguish catastrophic climate change governance from sub-catastrophic climate change governance as only those governance activities that reflect a nonlinear damage function.

To summarize, catastrophic and sub-catastrophic risks are often governed by the same assemblages of institutions, with significant overlap in the governance activities addressing both.

1.3 Research contributions and outline of the dissertation

This dissertation contributes to the literature on governance organizations that facilitate and formalize decision support relationships between experts and policy makers. This literature crosses a number of academic fields including science and technology studies, public policy, research policy, and organizational studies. While much of this literature concerns national governments, decision support is also an important process for sub-national governments, international and intergovernmental organizations, universities, think tanks, public research organizations, private companies, and other non-governmental organizations.

Specifically, this project explores decision support relationships at NASA and the Intergovernmental Panel on Climate Change (IPCC). These organizations are the primary knowledge brokers in the governance regimes addressing planetary defense and climate change respectively. NASA is active on both sides of the decision support equation – funding and conducting a significant proportion of global planetary defense R&D, while also serving as the focal institution for planetary defense decision-making. In contrast, the IPCC is an advisory organization that neither conducts independent research nor makes policy decisions. Rather, its mandate is to assess the state of climate change research conducted elsewhere and communicate that assessment to policy makers at national governments participating in the United Nations Framework Convention on Climate Change.

This dissertation is comprised of three research articles, which are followed by a short concluding chapter. Each article is briefly introduced. These introductions explain how each article fits into my larger dissertation project, how it contributes to existing scholarship, and my plans for publication and dissemination.

Article #1 begins by examining the contested, multidimensional concept of uncertainty itself. The paper contributes to the notoriously fragmented conceptual literature on uncertainty (Skinner et al. 2014), which can be generally understood as a discussion on the philosophy of science – but with most contributions coming from the interdisciplinary field of environmental risk. The paper presents a critical analysis of prominent typologies and conceptual frameworks that have emerged over the last 30 years. The paper proposes a series of amendments to the framework that has received the most scholarly attention to date (Walker et al. 2003; Kwakkel, Walker, and Marchau 2010) that result in a more comprehensive and epistemologically consistent conceptualization of uncertainty. This conceptual foundation forms the basis for my exploration of the decision support tasks of uncertainty reduction and communication in the proceeding articles.

Article #2 explores the process of systematic uncertainty reduction around planetary defense at NASA. This paper contributes to the R&D priority-setting literature that focuses on public research organizations – specifically, those that conduct what has been labelled "mission-oriented R&D" (Mowery 2012; Wallace and Ràfols 2015). Much of this literature approaches R&D priority-setting from a decision analytic perspective, highlighting the value of applying decision analytic methods to estimate the performance of R&D investments and improve decision-making (e.g. Keisler 2004; Bates et al. 2016; Barrett 2017; Drago and Ruggeri 2019; Bhattacharjya, Eidsvik, and Mukerji 2013; Arratia et al. 2016). A more recent strand of scholarship emphasizes how organizational and political factors also play a significant role in shaping research priorities (Brattström and Hellström 2019; Ciarli and Ràfols 2019; Hellström, Jacob, and Sjöö 2017; Cruz-Castro and Sanz-Menéndez 2018; D'Este et al. 2018; Wallace and Ràfols 2018, 2015). This article seeks to bridge the two perspectives by proposing a framework that shows how organizational and political factors are present at every stage of the "decision analysis cycle" – from initial problem definition to final project selection. This paper also constitutes the first descriptive study on planetary defense that draws on empirical data from expert interviews, bibliometric analysis, and a survey.

Article #3 investigates how uncertainties are communicated by scientists participating in the IPCC assessment process. This paper contributes to the critical literature on the IPCC's uncertainty language framework – a system designed to encourage the consistent characterization and communication of uncertainty in IPCC assessment reports. However, the recent literature on the uncertainty language framework focuses on how it was implemented during the IPCC's Fifth Assessment Report (AR5) cycle (2008-2014), with no consideration of the three special reports published in 2018 and 2019 (SR15, SRCCL, SROCC). These special reports constitute the latest application of the uncertainty language framework and perhaps provide the clearest indication of whether the issues raised since AR5 are being addressed in the AR6 cycle (2015-2021). Much of the existing literature also lacks a firm empirical basis, relying on the authors' own experiences implementing the framework and anecdotal examples drawn from the reports.

The article addresses both of these gaps in the literature by presenting a mixed methods analysis of the application of the uncertainty language framework by IPCC authors in recent special reports. A term usage analysis of the three SRs builds off of a similar analysis of AR4 and AR5 conducted by Mach et al. (2017) to examine language usage trends, while semi-structured interviews conducted with IPCC lead authors are used to examine these trends, as well as claims made in the literature about how language decisions are influenced by different interpretations of the framework. The fundamental aim of the paper is to reenergize the conversation around the communication of uncertainty in IPCC assessment reports and to this end, I propose a series of clarifications to the uncertainty communication guidance that can be applied in AR6 and beyond.

Both Article #2 and Article #3 contribute to the nascent literature on GCR governance. Much of the emerging GCR governance research has focused on imagining what effective GCR governance regimes might look like (Dafoe 2018a; Liu, Lauta, and Maas 2018; Millett and Snyder-Beattie 2017). However, there is a notable lack of scholarship mapping out what GCR governance looks like today. Two studies have applied decision analytic approaches to discuss how R&D might be prioritized to address the NEO impact hazard (Barrett 2017; Lee, Jones, and Chapman 2014) but neither describes the current state of planetary defense R&D priority-setting. It is my hope that the contributions made by this project to the almost non-existent descriptive literature on GCRs catalyzes a more vibrant discussion on how governance organizations are currently addressing GCRs – even if these efforts are inadequate given the magnitude and urgency of these threats.

The closing chapter revisits my key findings and discusses strategies for applying these insights at organizations tasked with either producing or using policy-relevant scientific knowledge.

Chapter 2

Article #1: Confidence, knowledge, and ignorance: Towards a coherent conceptualization of uncertainty

2.1 Preface

What is uncertainty? The central task of this article is to provide a comprehensive and useful answer to this deceptively simple question. When approaching this question, most people instinctively point to a lack of knowledge: if I am uncertain, then "I do not know" or "I do not know well." But several additional questions immediately emerge from this type of response. What does it mean to "not know" or to "not know well?" How does one measure degrees of uncertainty? Are there different types of uncertainty or just different degrees? This article contributes to a century-old conversation on these questions.

The other two articles in this dissertation each focus on one of the two core decision support tasks: uncertainty reduction and uncertainty communication. In the context of this dissertation, this article provides the conceptual foundation for understanding the nuances of those tasks. Notably, the article advances a "confidence-deficit" (i.e. Bayesian) interpretation of uncertainty that emphasizes the fundamentally subjective or intersubjective nature of estimations of how uncertain a given situation is and how easy it is to reduce.

This article builds off of a large intellectual endowment provided by the philosophical literature on the concept of uncertainty dating back to the work of Frank Knight nearly a century ago. Important contributions from the interdisciplinary field of environmental risk over the last 30 years have coalesced around a three-dimensional conceptualization of uncertainty that distinguishes between uncertainty *location*, uncertainty *level,* and the *nature* of uncertainty. I take this three-dimensional conceptualization as a starting point and advance the conversation by exposing and ironing out epistemological inconsistencies buried within conventional conceptualizations. I also apply useful concepts from complexity science like complex reflexive systems that connect the somewhat isolated conceptual literature on uncertainty to other fields intimately concerned with uncertainties emerging from system complexity.

I began exploring the concept of uncertainty in a major research paper entitled "Tail-dominance and catastrophe insurance: managing uncertainty in econometric climate change modeling," which I wrote in 2014 while completing a Master's degree at the Balsillie School of International Affairs. During that process, I first encountered the "W&H framework" presented in an article by Warren E. Walker and colleagues (Walker et al. 2003). I found the framework to be the most complete and practical typology of uncertainty in the literature. Five years on, the bibliometric analysis and literature review that I conduct in this article confirm that the framework is indeed the most influential and popular typology of uncertainty in the environmental risk literature to date. However, the framework has one rather surprising omission: the term *confidence*. As I argue in this article, confidence and subjectivist notions of probability have become the dominant paradigm for characterizing and communicating uncertainty in many disciplines and policy contexts today, including decision analysis and climate policy – two of the literatures explored in the other articles in this dissertation. Therefore, this article aims to address this "confidence gap" and to reenergize this conversation.

Most of the scholarship proposing frameworks, typologies, and taxonomies of uncertainty over the last 20 years have been published in journals of the authors' "home discipline," as opposed to interdisciplinary journals concerned more generally with the topics of risk and uncertainty. I plan to submit this article to one of these more interdisciplinary journals like *Risk Analysis* or *Journal of Risk Research*, which I believe are the more appropriate arena for this fundamentally interdisciplinary conversation.

2.2 Abstract

The fragmented conceptual literature on uncertainty has become increasingly standardized behind the tripartite distinction between uncertainty *location*, uncertainty *level*, and the *nature* of uncertainty popularized by the "W&H framework" (Walker et al. 2003; Kwakkel et al. 2010). However, the epistemological foundation on which the W&H framework is built is both vague and inconsistent. Perhaps most surprising is its avoidance of the term "confidence" – which has become the dominant perspective for characterizing and communicating uncertainty in many disciplines and policy contexts today. This article reinterprets the W&H framework from a Bayesian epistemological perspective, which understands uncertainty as a mental phenomenon arising from "confidence deficits" as opposed to the ill-defined notion of "knowledge deficits" that dominates the literature. This article proposes a number of amendments to the W&H framework, including a more consistent set of rules for determining when uncertainty may or may not be quantified, a clarification of the terms "ignorance" and "recognized

ignorance," and the expansion of the framework's level dimension to include levels of uncertainty reducibility. Lastly, this paper challenges the usefulness of the conventional distinction made between aleatory and epistemic uncertainty and proposes a more useful distinction based on developments in the field of complexity science that highlights the unique properties of complex reflexive (i.e. human) systems.

2.3 Introduction

2021 marks the 100th anniversary of Frank Knight's seminal book *Risk, Uncertainty and Profit* (1921), which has long served as the default citation for scholars differentiating between the slippery concepts of uncertainty and risk. However, a century later, the concept of uncertainty continues to frustrate philosophers, scientists, modelers, and policy makers alike. One of the principal reasons that uncertainty remains an unresolved topic is that it encompasses a multiplicity of overlapping concepts in addition to risk, such as ignorance, confidence, and ambiguity (Morgan and Henrion 1990). Over the last 40 years, there has been a proliferation of conceptual frameworks, typologies, and taxonomies of uncertainty attempting to comprehensively capture its key dimensions and create a common language for the characterization and communication of uncertainty by modelers and decision makers. In one systematic literature review, Skinner et al. (2014) identify 30 categorization frameworks for model-based policy analysis and decision support published since the early 1980s. Here, I identify an additional 22 (Appendix A). Despite this glut of uncertainty frameworks, one has been particularly influential: the so-called "W&H framework"[5] presented in Walker et al. (2003) and updated in Kwakkel et al. (2010).

The W&H framework is best known for popularizing the three-dimensional conceptualization of uncertainty that distinguishes between uncertainty *location*, uncertainty *level*, and the *nature* of uncertainty. Location refers to the different stages in the modeling and decision support process where uncertainty may arise; level refers to the degree to which something is uncertain; and nature describes the fundamental types of uncertainty. The W&H framework is, by far, the most cited framework in the transdisciplinary literature on uncertainty (see Section 2.4). Many of the purportedly novel frameworks published since the W&H framework have inherited its tripartite distinction between location, level, and nature or contain noticeable traces of its DNA (e.g. Refsgaard et al. 2007; Baustert et al. 2018; Mishra, Karmakar, and Kumar 2018; van der Keur et al. 2008). This harmonization in the conceptual literature on uncertainty over the last decade is evidence that the W&H framework has been enormously successful at

[5] The W&H framework is named after the first two authors of Walker et al. (2003) (Walker and Harremoës).

achieving its stated goal to integrate existing conceptualizations of uncertainty into a single overarching framework (Kwakkel et al. 2010).

However, one side effect of the homogenization of perspectives in the literature is a stagnation of critical commentary and innovation in the discourse. While the W&H framework has performed a crucial role in decluttering and focusing the conversation, the surprising absence of the term "confidence" in either Walker et al. or Kwakkel et al. reveals a disconnect between the W&H framework and the current discourse on model-based policy analysis and decision support. The conceptual apparatus associated with the concept of confidence—and by extension, the Bayesian interpretation of probability—has become the central paradigm for characterizing and communicating uncertainty in many policy domains. For instance, assessment reports produced by the IPCC, which are widely considered the gold standard for uncertainty communication, increasingly use confidence language to communicate uncertainties associated with knowledge claims (Janzwood 2020; Mach et al. 2017). Further, the Bayesian interpretation of probability is now the default perspective in fields such as decision theory, statistics, decision analysis, and uncertainty communication (Hill 2019).

Not only does the W&H framework not reflect or address the Bayesian perspective of uncertainty but it also does not meaningfully engage in a discussion of the underlying epistemological basis of the framework. In fact, this tendency to overlook basic questions on the nature of knowledge and how beliefs achieve the status of truth can be traced back all the way to Knight's original conceptualization of uncertainty and risk. While it may be justified to bypass epistemological quibbles in the name of pragmatism in certain situations, uncertainty is a foundational epistemological concept. And thus, a rigorous and transparent conceptualization of uncertainty must clearly articulate its epistemological axioms (Nearing et al. 2016).

I argue that conventional discussions around the W&H framework's dimension of uncertainty level—which is principally concerned with how to measure beliefs that fall somewhere between the poles of "true" and "false"—do not articulate a clear epistemological position. As a result, the popular framework entertains both frequentist and Bayesian interpretations of probability and thus proposes an inconsistent set of rules for determining when uncertainty may or may not be quantified. Additionally, the W&H framework conflates the distinct concepts of uncertainty and ignorance. I label this epistemologically vague or imprecise interpretation of uncertainty level exemplified by the W&H framework the *knowledge-deficit perspective*.

Acknowledging that “the overproduction of concepts signals a certain disarray” (Hettne 2005, pp. 544), this paper does not propose a completely new typology but instead reinterprets the W&H framework from a (moderate subjectivist) Bayesian perspective—or what could also be called a *confidence-deficit perspective*. From this perspective, uncertainty is understood as a mental phenomenon arising from “confidence deficits” in the quality of one’s knowledge as opposed to the imprecise notion of “knowledge deficits.”

This paper proposes several amendments to the W&H framework. Sidestepping the relatively straightforward dimension of uncertainty location, this article advocates a series of changes to the dimension of uncertainty level and the clarification of the terms “ignorance” and “recognized ignorance.” It also calls attention to the dimension of *uncertainty reducibility*. While the topic of uncertainty reduction is typically addressed in discussions about natures of uncertainty (where a distinction is made between fundamentally reducible and irreducible types of uncertainty), there is a surprising gap in the literature on the *extent* to which uncertainty can be reduced. I argue that uncertainty reducibility is a second way of distinguishing between greater and lesser uncertainties.

Drawing from debates in physics and philosophy on the limits of knowledge (Svozil 1996; Casti 1996; Popper 1982; Crutchfield et al. 1986), this paper also questions the usefulness of the conventional distinction between uncertainties arising from the natural variability of systems (aleatory uncertainty) and uncertainties arising from knowledge gaps (epistemic uncertainty), which are commonly presented as qualitatively distinct *natures* of uncertainty. While this distinction is deeply entrenched in the literature and can be found in nearly every uncertainty framework from the last 40 years, I argue that in most decision support contexts, one’s capacity to distinguish between aleatory uncertainty and epistemic uncertainty is, itself, deeply uncertain. Quite often, yesterday’s aleatory uncertainties turn out to be tomorrow’s epistemic uncertainties, as human knowledge evolves in unpredictable and nonlinear ways.

In most frameworks, the aleatory/epistemic distinction is used as a short-hand for communicating which uncertainties are believed to be reducible given the current state of knowledge and available resources and those that are not—a distinction that I suggest fits more comfortably within the dimension of uncertainty level. Rather than abandoning the dimension of nature altogether, I follow recent scholarship highlighting the uncertainties arising from language and incompatible knowledge frames (Brugnach et al. 2008; Dewulf et al. 2005; Ascough II et al. 2008; Beven 2016; Dewulf 2013; Refsgaard et al. 2013), and argue that these uncertainties, which arise from the uniquely reflexive nature of human systems, can serve as the basis for a more sound distinction between natures of uncertainty.

This article is organized as follows. Section 2.4 reviews the conceptual literature on uncertainty, focusing on the W&H framework. Section 2.5 critiques the W&H framework's dimension of uncertainty level and reinterprets it from a confidence-deficit perspective. This reinterpretation then informs an exploration of the related concepts of ignorance and uncertainty reducibility. Section 2.6 critiques the common aleatory/epistemic uncertainty distinction and draws from developments in the field of complexity science to advocate modified categories of uncertainty nature that highlight the unique properties of complex reflexive (i.e. human) systems. Section 2.7 summarizes these changes in an amended version of the W&H framework and Section 2.8 concludes with a discussion of the implications that these proposed amendments and reinterpretations hold for model-based policy analysis and decision support.

2.4 The W&H framework

The roots of the conceptual literature on uncertainty can be traced back to the early-to-mid-20th century, which saw important contributions from the fields of economics (Knight 1921; Keynes 1921), probability theory (de Finetti 1937, 1931; Ramsey 1926), and philosophy (Carnap 1950). Over the last 40 or so years, the mantle has been taken up by several disciplines concerned with model-based policy analysis and decision support, including, environmental science (Brugnach and Ingram 2012), water management (van der Keur et al. 2008), ecological assessment (Skinner et al. 2014), public health (Briggs, Sabel, and Lee 2009), and risk analysis (Morgan and Henrion 1990).

While some of the frameworks emerging from these literatures make a clear attempt to build off of previous scholarship, the evolution of the uncertainty concept has been disconnected and uneven. Some frameworks invent entirely new labels and apply them to old concepts, while other frameworks apply existing labels to new concepts. Take, for example, the term ambiguity, which is now commonly used to refer to uncertainties emerging from differences in beliefs, values, worldviews, and framings—but has alternately been used to describe uncertainties with poorly defined probabilities (e.g. Beer 2006; Dequech 2011). However, since the publication of the W&H framework in Walker et al. (2003), which was amended by Kwakkel et al. (2010), there has been a fairly rapid harmonization of the conceptual language used to describe uncertainty in the literature.

According to Google Scholar, the W&H framework presented by Walker et al. is, by far, the most cited uncertainty categorization framework in the literature with 1,958 citations.[6] The update in Kwakkel et al. contributes an additional 150 citations. An analysis of the year-to-year citation record of the W&H framework shows that citations have increased steadily since its publication, peaking in 2019 when the framework was cited more than twice as many times as 2010 (Fig. 1). The enduring influence of the W&H framework illustrates that many scholars addressing challenges associated with computational modeling or model-based decision support have gravitated towards the tripartite conceptualization of uncertainty, which divides uncertainty along three dimensions: location, level, and nature. These three dimensions also resonate amongst scholars who have developed derivative frameworks that aim to build off of or even replace the W&H framework (e.g. Dewulf et al. 2005; Mishra, Karmakar, and Kumar 2018; Refsgaard et al. 2007; Skinner et al. 2014).

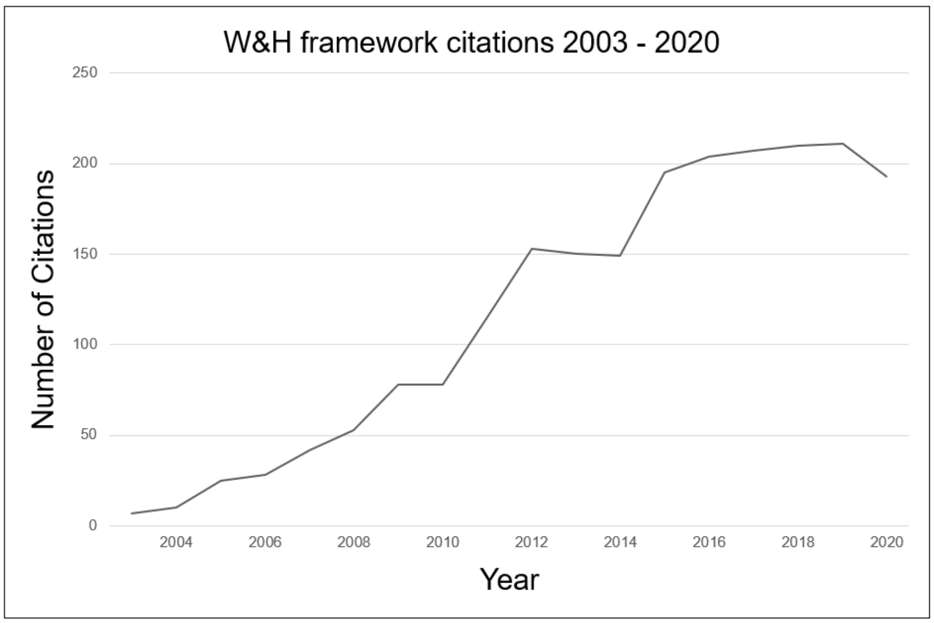

Figure 1: Combined year-to-year citations of Walker et al. (2003) and Kwakkel et al. (2010) (source: Google Scholar)

[6] (As of December 31, 2020). I conducted a systematic review of the literature, using the SCOPUS database that added 22 frameworks to the 30 frameworks identified by Skinner et al. (2014). I then compared how many times all 52 frameworks were cited using data from Google Scholar. One source (Morgan and Henrion 1990), was cited more than the combined citations of Walker et al. (2003) and Kwakkel et al. (2010). However, Morgan and Henrion (1990) is widely considered a seminal book in the fields of model-based policy analysis and risk analysis with wide-ranging contributions to the literature. However, its uncertainty taxonomy has not had nearly the influence of the W&H framework.

Walker et al. define uncertainty as a departure from (the unachievable ideal of) "perfect knowledge." Citing van Asselt (2000), the authors add that "Uncertainty is not simply a lack of knowledge, since an increase in knowledge might lead to an increase of knowledge about things we do not know and thus increase uncertainty" (Kwakkel et al. 2010, pp. 301).[7] Even if knowledge gains do not necessarily chart a linear path towards perfect knowledge, this definition implies that the accumulation of *all* knowledge will eventually lead you along the asymptote towards a state approaching perfect knowledge. As such, the W&H framework can be said to reflect a knowledge-deficit interpretation of uncertainty. Another knowledge-deficit interpretation is made by Dewulf et al. (2005) who define uncertainty as a lack of knowledge or information about a phenomenon.

Of the three dimensions of uncertainty, *location* is the most straightforward. Location maps the points where uncertainty emerges as models are conceived, built, tested, and their outputs communicated. While models are built to manage or resolve uncertainties, they also inject entirely new uncertainties into the decision context—uncertainties that arise from the modeling process itself. In any given modeling study, key uncertainties may arise from: (1) bounding the system; (2) the conceptual model; (3) the computer model (model structure and parameters); (4) data inputs; (5) implementing the model; or (6) communicating model outputs (Kwakkel et al. 2010).

The dimension of uncertainty *level* attempts to define the characteristics that distinguish greater uncertainties from lesser uncertainties—or what makes something more uncertain than something else. Walker et al. describe uncertainty level as a continuum from the unachievable ideal of complete determinism to total ignorance—or alternatively, from "know" to "no-know." The W&H framework identifies four levels of uncertainty: shallow uncertainty, medium uncertainty, deep uncertainty, and recognized ignorance.[8] The levels are distinguished from one another according to the extent to which all possible outcomes are known and the extent to which the uncertainty can be described probabilistically.

Lastly, the dimension of *nature* addresses whether there are qualitatively distinct *kinds* of uncertainty that can help guide the decision-making of modelers and individuals using model outputs. The W&H

[7] Take, for example, a decision maker who is uncertain about the outcome of rolling a six-sided die. They may assume that they have a roughly one in six chance of correctly guessing the outcome (based on the mathematical symmetry of the die) but suspect that the die is not a perfect cube. Suppose the decision maker decides to try to reduce uncertainty about the outcome by increasing their knowledge about the physical properties of the die and they discover that the manufacturer has a reputation for secretly "loading" its dice. By acquiring this new piece of knowledge, the decision maker is actually more uncertain about the outcome of rolling the die.

[8] Walker et al. (2003) only identify three levels of uncertainty and refer to them by different names. However, here I refer to the revised four-level version of the framework presented in Kwakkel et al. (2010).

framework distinguishes between ontological uncertainty (which is more commonly called aleatory uncertainty) and epistemic uncertainty. Aleatory uncertainty describes the inherent variability associated with real-world systems, which cannot be reduced through learning or gathering evidence, while epistemic uncertainty describes uncertainty arising from knowledge deficits that can hypothetically be reduced or eliminated. Acknowledging contributions made in the literature since the publication of Walker et al. on uncertainties arising from human language, beliefs, values, and knowledge frames, Kwakkel et al. expand the nature dimension to include a third category they label ambiguity.

While location addresses the question "*where* is uncertainty?" level and nature collectively dive deeper into the question "*what* is uncertainty?" The following section begins with an examination of how the W&H framework (and similar frameworks reflecting a knowledge-deficit interpretation of uncertainty) conceptualize the dimension of uncertainty level. I then make the case that the knowledge-deficit perspective is limited by its failure to clearly articulate a consistent epistemological position on the nature of knowledge and how we come to "know" it.

2.5 Level of uncertainty

At its core, the W&H framework's notion of uncertainty level is an attempt to describe the *degree* of uncertainty and to define the characteristics that separate greater uncertainties from lesser uncertainties. For instance, we know intuitively that the weather tomorrow in a particular location is less uncertain than the weather a month from now. But it is surprisingly difficult to specify *how much* more uncertain the latter is than the former or precisely what characteristics make one situation more uncertain than another. I use the term "uncertainty situation" broadly to describe instances where a decision maker confronts uncertainty about the past, present, or future.

Level is the most difficult dimension of uncertainty to characterize but it is also the component of uncertainty that decision makers care about most. Decision makers demand absolute measurements of uncertainty level, such as the precise probability values of plausible outcomes, in order to make expected cost (or expected value) calculations and conduct cost-benefit analyses. Even when absolute measurements of uncertainty are unavailable, relative measurements such as rankings can help decision makers identify the largest gaps in their understanding and allocate resources accordingly. But what exactly does it mean for someone to be more or less uncertain about something?

There are two basic ways to approach this question. First, we might say that it is *more difficult to predict* the outcome of a more uncertain situation than the outcome of a less uncertain situation. For example, it seems intuitive that it is easier to predict the outcome of a coin flip than predicting the exact date of the next pandemic outbreak. This interpretation is adopted by most uncertainty frameworks, including the W&H framework. Described by Bradley and Drechsler: "the [level] dimension relates to the *difficulty* the agent has in making a judgement about the prospects they face" (2013, pp. 7, emphasis in original). From this perspective, the central task for characterizing uncertainty level is identifying the properties that make outcomes easier or harder to predict.

Second, we might say that uncertainty is *more difficult to reduce* for a more uncertain situation than a less uncertain situation. For example, uncertainty around whether a coin is rigged can be reduced fairly easily by analyzing the composition of the coin or conducting a sufficient number of trials, whereas uncertainty around the timing of a disease outbreak is significantly more difficult to reduce. The predictability and reducibility approaches are linked. Prediction usually — but not always[9] — becomes easier when uncertainty is reduced. First, I address the prediction interpretation of uncertainty level (Section 2.5.1) and the contentious concept of ignorance (Section 2.5.2), before returning to the issue of reducibility (Section 2.5.3).

2.5.1 Predictability

A common starting point for differentiating between levels of uncertainty in many frameworks is Knight's classic definitions of risk and uncertainty (1921). Today, "Knightian risk" and "Knightian uncertainty" are commonly used as a shorthand for quantifiable and unquantifiable uncertainty respectively (Sigel et al. 2010; Kwakkel et al. 2010; Chen et al. 2007; Beer 2006). According to this simple framework, quantifiable uncertainties are less uncertain (that is, the decision maker is closer to perfect knowledge) than unquantifiable uncertainties. This distinction provides a simple criterion for determining how difficult it is to predict an outcome: outcomes are easier to predict when uncertainty can be quantified than when it cannot. However, this criterion is of little use if we do not know when it is and is not appropriate to quantify uncertainty.

Fortunately, most frameworks, including the W&H framework, attempt to define the characteristics that allow uncertainty to be expressed probabilistically (i.e. quantified) in some situations but not in others. The W&H framework distinguishes between four levels of uncertainty (shallow uncertainty,

[9] Sigel et al. (2010) point out that the generation of new knowledge (i.e., the reduction of uncertainty) can sometimes unearth new uncertainties making prediction even more difficult. Oppenheimer, et al. (2008) call the phenomenon of information acquisition that increases the divergence between current beliefs and the actual outcome "negative learning."

medium uncertainty, deep uncertainty, and recognized ignorance) (Fig. 2). Others have assigned these levels different labels like "statistical uncertainty" (i.e. shallow uncertainty), "scenario uncertainty" (i.e. deep uncertainty), and "qualitative uncertainty" (i.e. recognized ignorance) (Brouwer and De Blois 2008; Refsgaard et al. 2007; van der Keur et al. 2008; Knol et al. 2009). Each level is defined by two criteria: (1) whether all possible outcomes can be identified, and (2) whether probability values can be assigned to each outcome.

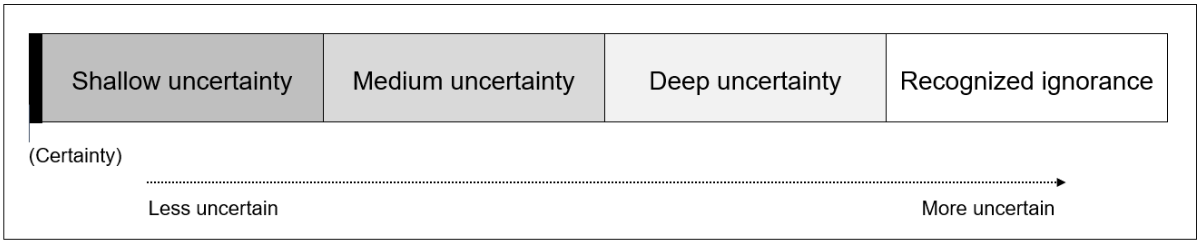

Figure 2: Levels of uncertainty (adapted from Kwakkel et al. 2010)

Shallow uncertainty is one step away from perfect knowledge, describing situations where all possible outcomes are known and the distribution of probability across those outcomes is also known. For instance, when rolling a six-sided die in a controlled environment, all possible outcomes are known (sides one to six) and the probability for each outcome is also known (~16.67 percent). Shallow uncertainties can be precisely quantified using probability values. The term medium uncertainty is used to describe a subset of shallow uncertainties that can be ranked relative to one another but not precisely quantified (i.e. fuzzy sets). Deep uncertainty is used to describe situations where all possible outcomes are known but the precise distribution of probability across those outcomes is unknown. An example of deep uncertainty is rolling a randomly "loaded" six-sided die where you know the die is loaded but not precisely *how* it is loaded. Therefore, the skewness of the probability distribution is unknown and you cannot allocate probabilities to all possible outcomes.

The furthest level from perfect knowledge is what the W&H framework refers to as recognized ignorance. Recognized ignorance describes situations where all possible outcomes cannot be identified, which implies that the distribution of probability across outcomes is also unknown. For example, we could imagine rolling a die with an unknown number of sides. In this situation, we are unable to confidently assign probability values across the unknown number of outcomes whether the die is loaded or not.

These two criteria — the identifiability of outcomes and the ability to confidently assign probabilities — provide a straightforward way to distinguish between situations where predictions are easier to make and situations where predictions are harder to make. Recent contributions to the literature seem to overlook the fact that Knight originally proposed these same two criteria back in *Risk, Uncertainty, and Profit* (Langlois and Cosgel 1993). Knight writes: "particular occurrences [are] foreseeable, if only all the alternative possibilities are known and the probability of the occurrence of each can be accurately ascertained." (1921, pp. 199). Combining these two criteria, each level can be said to possess a specific *uncertainty structure* that describes the extent to which we are able to accurately define a probability distribution function (Table 2).

Table 2: "Uncertainty structure" of uncertainty levels

Level	Uncertainty Structure		Measurement scale	Example
	All possible outcomes identified	Probabilities can be assigned		
Shallow uncertainty	Yes	Yes	Ratio / interval / ordinal	Six-sided die
Deep uncertainty	Yes	No	Nominal / categorical	Loaded six-sided die
Recognized ignorance (Knightian uncertainty)	No	No	N/A	*X*-sided die

Kwakkel et al. suggest that each uncertainty level coincides with specific scales of measurement (Table 2). Shallow uncertainties can either be described using precise probabilities (ratio or interval scale) or simply ranked without information on the magnitude of the difference between the likelihood of possible outcomes (ordinal scale). For deep uncertainties, possible outcomes can only be defined according to a nominal or categorical scale, while recognized ignorance has no corresponding measurement scale.

Uncertainty structure provides a sensible framework for defining a continuum from lower uncertainty to higher uncertainty. According to the W&H framework, knowledge deficits are the most significant for uncertainty situations characterized by recognized ignorance and least significant for situations characterized by shallow uncertainty. However, important questions remain unanswered. How do we know when it is appropriate to assign probability values to possible outcomes and when it is not? And how do we know if we have identified all possible outcomes? In other words, how do we go about sorting real-world uncertainties into these levels? These questions reveal two ingredients missing from frameworks based on a knowledge-deficit interpretation of uncertainty: a consistent and scientifically rigorous understanding of probability and a corresponding theory of knowledge.

2.5.1.1 Probability theory and epistemology

Absent from the W&H framework—and most of the literature that professes to offer a comprehensive and useful framework for characterizing and communicating uncertainty—is any significant discussion of the epistemological approach that underpins the framework. That is not to say that these frameworks lack an epistemology, as it would be impossible for a discussion of different levels of uncertainty to not appeal to some set of criteria for determining when it is possible or appropriate to assign probability values to possible outcomes and when it is not—even if these criteria are not stated explicitly.

Frameworks reflecting a knowledge-deficit interpretation of uncertainty tend to avoid clearly articulating these criteria. For instance, Walker et al. define shallow uncertainty as "any uncertainty that can be *described adequately* in statistical terms" (2003, pp. 12, emphasis added) but do not provide criteria for what would constitute an "adequate" statistical description. Meanwhile, Kwakkel et al. hint that shallow uncertainties can be described by either objective (frequentist) or subjective (Bayesian) probabilities but do not elaborate on how these two types of probabilities are generated—nor do they investigate whether those two understandings of probability can coexist within a single uncertainty framework. A practical conceptualization of uncertainty level that can be applied by modelers and decision makers to describe real-world uncertainties should clearly articulate when statistical description is appropriate and when it is not.

Knight (1921) proposes one set of rules for determining when it is appropriate to assign probability values to possible outcomes. Like Kwakkel et al., Knight entertains both the objectivist and subjectivist notions of probability. For Knight, two criteria must be met for an uncertainty situation to be described probabilistically: (1) the existence of a large set of historical instances or "trials" and (2) a high degree of similarity between historical instances, or "trial homogeneity." Therefore, for Knight, adequate statistical description means being able to apply frequentist methods to generate objective probabilities. However, like other knowledge-deficit conceptualizations of uncertainty level (e.g. Baecher and Christian 2000; Stirling 1998), Knight does not fully commit to an entirely frequentist perspective of probability, switching to the Bayesian perspective when frequentist methods become untenable. This “flip” is illustrated in Knight’s description of three types of uncertainty situations.

First, Knight describes situations where probability can be measured from the mathematical symmetry of the experiment, such as a coin flip under controlled conditions. This perspective sees probabilities as innate tendencies or “propensities” that can be derived from the physical characteristics of the system (Climenhaga 2019). However, according to Knight, almost no real-world uncertainties are characterized by such mathematically derived propensities.

Second, adopting a frequentist understanding of probability, he describes relatively rare situations where historical frequencies can be used to quantify the likelihood of future outcomes (which he calls “risks”). The key characteristic of these situations is that trials are “more or less homogeneous” (Runde 1998, pp. 543). In other words, with a sufficiently large number of nearly identical trials, the rate at which something occurred in the past can be expressed probabilistically and serve as an accurate indicator of how likely it is to happen in the future. For example, the rate of babies born female, male, or intersex in the past can be used to express the likelihood of a baby being born with a particular sex in the future.

However, as Knight warns, the frequentist approach falters when trials are insufficiently homogeneous or if there is a lack of historical data. For instance, when predicting the outcome of unique or unprecedented events, it is unclear which events would constitute a relevant population of trials (Morgan and Henrion 1990). Knight’s third uncertainty situation, which he refers to as “estimates” describes situations where “there is no valid basis of any kind for classifying [trials]” (1921, pp. 224). Knight argues that estimates constitute the vast majority of real-world uncertainty situations. Due to the untenability of frequentist methods for calculating probabilities, estimates can only be described qualitatively using subjective judgment. In other words, when frequentist calculations are not possible, one must discard the frequentist notion of probability and adopt something akin to a Bayesian perspective of probability as a subjective degree of belief or confidence (even though Knight’s writing preceded much of the pioneering work in the field of Bayesian statistics by Keynes, de Finetti, Ramsey, Savage, and others).

It is not entirely clear how Knight’s rules might apply to the W&H framework’s levels of shallow uncertainty, medium uncertainty, deep uncertainty, and recognized ignorance. It seems safe to assume that situations where probabilities can be generated by rigorous frequentist methods would align with the level of shallow uncertainty. However, it is less clear whether shallow uncertainties can also be quantified using Bayesian methods.

To illustrate the difference between the Bayesian and frequentist perspectives on probability, consider the statement “there is a 70 percent probability that Candidate A will win the election.” From a frequentist perspective, a 70 percent probability is strictly empirical—since 70 percent of trials (e.g. “runs” of an election model simulation) resulted in Candidate A winning the election, then approximately 70 percent of identical future trials will result in Candidate A winning. Often, the frequentist perspective is paired with a “possible worlds” perspective where a 70 percent probability is interpreted as “seven out of ten worlds where Candidate A wins” and “three out of ten worlds where Candidate A loses.” From a Bayesian perspective, a 70 percent probability means that after considering all evidence (which might include many “runs” of an election model simulation) the decision maker is “70 percent confident” that Candidate A will win tomorrow.

Knight is not alone in suggesting that Bayesian probabilities have a lower “epistemic value” than frequencies. For example, Stirling (1998, pp. 102) prescribes using frequentist methods when there is a “firm basis for probabilities” and Bayesian methods when there is not. According to this perspective, quantification is more appropriate when uncertainties can be expressed using frequentist probabilities and less appropriate when they are derived from other evidence and incorporated into a decision maker’s beliefs using Bayesian methods. However, Bayesians would not only disagree that subjective probabilities are somehow “valid but inferior” but would also argue that acknowledging the fundamentally subjective nature of probability is the only epistemologically defensible way to characterize uncertainty level.

The frequentist and Bayesian perspectives reflect different theories of knowledge that disagree on how a belief achieves the status of knowledge (i.e. justified, “true” belief). The frequentist approach advocates a statistical procedure (calculating frequencies) that is believed to outperform other procedures at making accurate predictions in the long run. Therefore, the frequentist approach is most closely aligned with the reliabilist program of epistemology (Woodward 1998). For reliabilists, it is more important to have reliable processes that frequently yield truths than to chase after purportedly infallible processes (Goldman 1998). Therefore, according to the frequentist perspective, a belief that the rate of historical outcomes reflects the likelihood of future outcomes is considered “true” because it is generated by a reliable procedure.

Bayesian epistemology, on the other hand, is more dogmatic in its commitment to the view that probability is inextricably linked with the mental processes of individuals. For Bayesians, these subjective beliefs of individuals are justified by repeatedly updating one's priors using Bayes' theorem, thereby approaching (but never reaching) the status of "truth" (Hartmann and Sprenger 2011). While Bayesians disagree amongst themselves about the extent to which subjective beliefs should be reined-in by rational constraints (e.g. coherence with deductive logic), they are bound by the view that probability reflects one's degree of belief or confidence in the truthfulness of a claim.

For a frequentist, likelihood (and uncertainty) can be quantified when there is a sufficient number of homogenous historical trials — despite the fact that there are no clear rules for determining what constitutes a sufficient number of trials and it is often difficult to determine the relevant criteria for determining trial homogeneity. For a Bayesian, uncertainty can, in theory, always be quantified—even if their confidence in the extent or quality of their knowledge is very low. However, in many cases when a decision maker's confidence is low, they will forgo the use of probability values and instead rely on other techniques for measuring and communicating uncertainty level like probability ranges, ranking possible outcomes, or providing qualitative descriptions of uncertainty level.

The rule set proposed by Knight entertains both the frequentist and Bayesian interpretations of probability, which are based on different understandings of what knowledge is and how we come to "know" it (frequencies versus beliefs). However, there is disagreement about whether the frequentist and Bayesian approaches can exist side-by-side within one conceptualization of uncertainty (like Knight's conceptualization). Baecher and Christian (2000) argue that frequency and belief are not necessarily incompatible and may address uncertainties in "different realities." However, I agree with Nearing et al. (2016) who contend that a logically coherent conceptualization of uncertainty requires a consistent and transparent epistemology.

I have two reasons for advocating a confidence-deficit (Bayesian) interpretation of uncertainty. First, the Bayesian epistemological perspective has emerged as the dominant paradigm in most scientific fields today (Morgan 2014).[10] A useful conceptual framework of uncertainty needs to resonate with the modelers and decision makers that use it. Second, a confidence-deficit perspective does not preclude the use of frequentist methods to inform likelihood assessments. From a Bayesian perspective, frequencies may still have epistemic value insofar as they may help inform a prior probability distribution — that is, a

[10] Perhaps with the exception of the health sciences (Morgan 2014).

decision maker *believes* that the past behaviour of a system is a good indicator of its future behaviour. In fact, repetition is a common way to establish or update one's priors (Sigel et al. 2010). Explained by Granger Morgan: "when large quantities of evidence are available on identical repeated events, one's subjective probability should converge to the classical frequentist interpretation of probability" (2014, pp. 7176). Therefore, the number and homogeneity of trials may constitute useful *signals* that affect a decision maker's confidence in the accuracy of their prior probability distribution.

From this perspective, frequentist calculations of probability are simply a tool that can be used to inform subjective judgements. When a decision maker assigns a probability value to an uncertain outcome that was generated from the frequency of historical trials, it means that they have a high degree of confidence that the historical trials were sufficiently similar to one another and sufficiently similar to the next trial (the future outcome). For example, a 50 percent probability that a coin will land on "heads" calculated from 100 trials is the same as saying that an individual has a high degree of confidence that the 101st trial is sufficiently similar to the preceding 100 trials.

The starting point for a Bayesian or confidence-deficit interpretation of uncertainty level is the view that probability reflects a decision maker's degree of belief or confidence level in the "truthfulness" of a prediction. A knowledge-deficit interpretation of uncertainty distinguishes between various levels of uncertainty depending on the extent to which the decision maker *knows* all possible outcomes and their relative likelihood. From a confidence-deficit perspective, levels of uncertainty can be distinguished by the extent to which the decision maker is *confident that they know* all possible outcomes and the relative likelihood of outcomes.[11]

A few definitions of uncertainty in the literature reflect the confidence-deficit perspective. For instance, uncertainty has been defined as a lack of confidence about the specific outcomes of an event (Refsgaard et al. 2007) and as a lack of confidence about one's knowledge relating to a specific question (Sigel et al. 2010). However, no intellectual community has embraced the confidence-deficit perspective as fully as the field of decision analysis, with one popular introductory textbook remarking: "Because there is no

[11] Good (1971) notes that there are nearly as many forms of Bayesianism as there are Bayesians. Therefore, it is worth specifying that I adopt a *moderate subjectivist* interpretation of Bayesian epistemology that permits subjective beliefs to be reined-in by rational constraints (e.g., coherence with deductive logic) and, following Joyce (2005), allows subjective degrees of belief to reflect both the "balance" and "weight" of evidence. While Bayesianism has been criticized for inadequately accounting for the weight or quantity of evidence (Keynes 1921), some Bayesian theorists like Joyce (2005) argue that subjective judgements can be updated to reflect both the balance or direction that evidence is pointing, as well as the amount of evidence (see: Hill 2019).

such thing as an objective probability, using a term like 'subjective probability' only creates confusion" (Howard 2007, pp. 35).

Here, I define uncertainty as: *a mental state of imperfect confidence in the extent or quality of one's knowledge about the outcomes of an event*. Imperfect confidence is *any level of the perceived truthfulness of a claim that can be increased*.

Conventionally, shallow uncertainty has been defined as situations where all possible outcomes are known and the distribution of probability across those outcomes is also known. Through a Bayesian lens, it is reinterpreted as: situations where one has *high confidence* that all possible outcomes are known and *high confidence* in the relative likelihood of those outcomes. Similarly, deep uncertainty is reformulated as: situations where one has *high confidence* that all possible outcomes are known but *lacks confidence* in the relative likelihood of those outcomes.

Lastly, recognized ignorance describes situations where one *lacks confidence* that all possible outcomes have been identified and *lacks confidence* in the relative likelihood of those outcomes. What makes this level of uncertainty particularly interesting is that by lacking confidence that all possible outcomes have been identified, it is necessarily the case that we can imagine other possible outcomes, or at least have some notion of what lies beyond the boundary of our knowledge. However, a few scholars have been quick to point out that the term recognized ignorance is, in fact, an oxymoron.

2.5.2 Ignorance

A conceptualization of uncertainty level must also address the related but distinct concept of ignorance. Unrecognized ignorance (i.e. "unknown unknowns") describes a "void in our knowledge," while recognized ignorance (i.e. "known unknowns") describes "knowledge of a void" where we are not only aware of the limits of our knowledge, but we have a sense of what it is we do not know. But as Wynne (1992) and Brown (2004) point out, "knowledge of a void" does not really constitute ignorance at all, which, by definition, escapes recognition.

While uncertainty is a function of the confidence level of a decision maker, true (unrecognized) ignorance describes a knowledge void that is completely outside of a decision maker's awareness (e.g. a die has a secret seventh side), and thus cannot factor into a decision maker's assessment of their confidence. True ignorance is external—it has nothing to do with the mental processes of a decision maker and can intervene even when confidence in the uncertainty structure is high. In contrast, the identification of known unknowns relies on mental models produced by the decision maker and subjective judgements of the accuracy of those models. In other words, known unknowns are a dimension of uncertainty, not ignorance.

In situations involving known unknowns, an individual possesses a rough model of the pieces of evidence that, when combined, would give them a high degree of confidence in an outcome. For example, when predicting the outcome of rolling a die, a decision maker would want to know the number of sides and the mechanics of how it is rolled. A known unknown exists when a decision maker believes that something is missing from this model (e.g. they believe that it might be possible for the die to land on an edge or vertex). Therefore, known unknowns are hypothesized knowledge gaps in our mental models of the uncertainty situation. And as such, they are characterized by uncertainty themselves.

Reformulated from a Bayesian perspective, known unknowns describe situations where a decision maker has some degree of confidence that: (1) their mental model of the uncertainty situation is generally accurate but in some way incomplete or inaccurate, and (2) they have some sense of what is missing from the model or which component is inaccurate.

An historical example is the 1964 prediction of the existence of the Higgs boson, which physicists hoped would resolve the uncertainty around the Standard Model of particle physics and the proposed Higgs field. The model hinged on an unproven hypothesis: the existence of the Higgs boson. In this case, the physicists hoped to increase their confidence in the accuracy of their model by obtaining conclusive empirical proof of the hypothesized elementary particle, which was eventually discovered in 2012 during a series of experiments using the Large Hadron Collider at CERN. Therefore, back in 1964, the existence of the Higgs boson constituted a known unknown.

Many conceptualizations of uncertainty level, including the W&H framework, use the term recognized ignorance to describe the uncertainty level furthest from certainty (e.g. Walker et al. 2003; Petersen et al. 2003; Petersen 2012; Refsgaard et al. 2007, 2013; van der Keur et al. 2008; Brouwer and De Blois 2008; Knol et al. 2009; Mishra, Karmakar, and Kumar 2018). However, known unknowns do not only emerge in situations where the uncertainty structure is haziest. They exist at every level. When dealing with shallow uncertainty, we may be able to describe a missing piece of knowledge that would allow us to go from quantifying the probabilities of different possible outcomes to predicting the outcome with near-perfect confidence. When facing deep uncertainty, we may be aware of precisely the information we are missing that would allow us to confidently assign probability values. And for situations where we are unable to confidently identify all possible outcomes, we may be able to outline the void in our understanding that—if filled—would allow us to confidently identify all possible outcomes.

To eliminate confusion around the term “recognized ignorance” (which has nothing to do with ignorance), I prefer to call this deepest level of uncertainty “Knightian uncertainty,” recognizing Knight as the first person to highlight the uncertainty structure where decision makers are unable to identify all possible outcomes. The relationship between ignorance, known unknowns, and uncertainty level is illustrated in Figure 3.

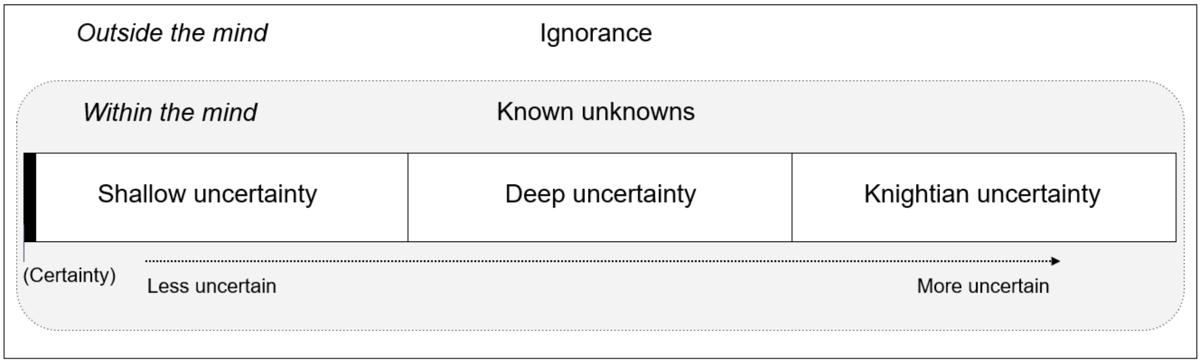

Figure 3: Relationship between ignorance, known unknowns and uncertainty level

2.5.3 Reducibility

So far, I have examined the first of two interpretations of what it means for one uncertainty situation to be more uncertain than another—that it is harder to predict a more uncertain outcome than a less uncertain outcome. A second way to interpret uncertainty level is that greater uncertainties are harder to reduce than lesser uncertainties.

Most frameworks tend to address reducibility when differentiating between natures of uncertainty, not levels. Typically, a distinction is made between fundamentally reducible and fundamentally irreducible uncertainties (see Section 2.6). However, for uncertainties that *can* be reduced, there has been surprisingly little consideration of the relative difficulty of reducing different uncertainties, with the exception of brief discussions in Matott et al. (2009) and Sigel et al. (2010). Intuitively, we know that some uncertainties are more reducible than others. For example, it is easier to reduce uncertainty about whether or not it will rain tomorrow (I could simply check the weather forecast) than whether or not it will rain a month from now.

If uncertainty is a state where we lack perfect confidence in the extent or quality of our knowledge, then a reduction of uncertainty is a step towards perfect confidence. To reduce uncertainty is to increase one’s confidence level that all possible outcomes have been identified and probability has been accurately distributed amongst possible outcomes. While uncertainty structure provides a framework for distinguishing between the properties that define states of low confidence, intermediate confidence, and high confidence, uncertainty reducibility is concerned with predicting potential *increases* in confidence.

However, reducibility is not only concerned with the extent that confidence can increase, but how easily those increases can be achieved—that is, the increase in confidence *per unit of effort at the margin*. More specifically, uncertainty reducibility is concerned with *predicting* confidence increases per unit of marginal effort *ex ante* (rather than simply measuring confidence increases per unit of marginal effort after the fact).

Therefore, proposing that some uncertainties are more reducible than others is the same as suggesting that the confidence level of a decision maker is more easily increased (i.e. it requires less effort) for some uncertainties than for others. The Bayesian perspective provides a clear-cut procedure for increasing confidence: acquiring new evidence and repeatedly applying Bayes' formula to incorporate that evidence into the decision maker's subjective assessment.

To estimate how reducible a particular uncertainty is, a decision maker constructs an evidence model describing the evidence that would be necessary to significantly increase their confidence in the extent or quality of their knowledge. Relevant evidence can come in many forms, including new observations, trials, and opinions elicited from experts. For example, when predicting the outcome of rolling a die, the decision maker can increase their confidence by learning how many sides the die has and the mechanics of how it is rolled. Supporting evidence might include: a count of the number of sides, an x-ray revealing potential imperfections, and the results of 1,000 rolls. Evidence can also come in the form of new theories or concepts that illuminate relevant system dynamics and increase the decision maker's confidence that the model is an accurate reflection of the real-world system.

Implicit in these evidence models is a prediction of how much the decision maker's confidence will increase if all the evidence is successfully gathered. As noted previously, measuring increases in confidence is ultimately a subjective or intersubjective process and can be expressed probabilistically. Increases in a decision maker's confidence (i.e. reductions in uncertainty) can be "pegged" to one of the three uncertainty structures identified in Section 2.5.1 (shallow uncertainty, deep uncertainty, and Knightian uncertainty). The decision maker who was previously in a state of Knightian uncertainty may predict that when a particular set of evidence is gathered, their confidence in the possible outcomes and the distribution of probability across those outcomes will increase, thereby pushing them into a state of deep or shallow uncertainty.

The next step is for a decision maker to estimate the amount of effort necessary to collect the evidence that would increase their confidence. I refer to this process as the construction of the "evidence-effort model." Effort is a catch-all term for investments necessary to uncover new evidence and increase a decision maker's confidence. These investments can include time, money, physical energy, and computational resources, as well as less tangible costs such as organization, ingenuity, and cognitive capacity. Individually, none of these indicators seems to capture the entire investment necessary to reduce most uncertainties that decision makers are concerned about. For example, a focus on the financial costs of uncertainty-reducing research and development (R&D) fails to account for the fact that money is spent more efficiently and effectively by some organizations than others (MacAskill 2015). Attempts to integrate different types of effort into a single measurement are inherently subjective (or intersubjective).

Here, I identify two broad categories of uncertainty reducibility: uncertainties that are practically reducible and uncertainties that are practically irreducible. Practically reducible uncertainty is defined as situations where imperfect confidence in the extent or quality of one's knowledge regarding an outcome can be increased (given available time, money, physical energy, organization, ingenuity and cognitive resources) to a specified level. The specified level depends on the decision context. For example, if a decision maker requires precise probability values for every plausible outcome, then a practically reducible uncertainty situation is one where it is possible to gather evidence that would result in a high level of confidence in the relative likelihood of each outcome (i.e. shallow uncertainty). Practically irreducible uncertainty describes situations where the effort requirements to achieve a specified confidence level exceed what is available.[12]

While the W&H framework does not address degrees of uncertainty reducibility, it does use the concept of uncertainty reducibility to distinguish between two fundamentally distinct uncertainty "natures."

2.6 Nature of uncertainty

While level considers uncertainty as a matter of *extent*, nature considers uncertainty as a matter of *kind*. Typically, uncertainty is sorted into two distinct natures: epistemic uncertainty, which is fundamentally reducible, and aleatory uncertainty, which is fundamentally irreducible. By proposing two fundamentally different natures of uncertainty, existing frameworks make two implicit claims.

[12] The decision to classify an uncertainty situation as practically reducible or irreducible may itself be highly uncertain, creating a separate uncertainty situation that itself may or may not be practically reducible.

First, they assert that aleatory and epistemic uncertainty constitute qualitatively distinct kinds of uncertainty. And second, as most of these conceptualizations are presented as being *practical* frameworks for real-world decision support contexts (e.g. Baecher and Christian 2000; Refsgaard et al. 2007; Skinner et al. 2014; Walker et al. 2003; Kwakkel, Walker, and Marchau 2010), they also imply that distinguishing between natures provides analytical value for modelers and decision makers beyond the existing distinctions made between levels and locations. However, I argue that neither claim is particularly strong. Here, I address the problems with the aleatory/epistemic distinction before proposing a new, more useful framework for distinguishing between different natures of uncertainty.

2.6.1 **(In)determinism**

Aleatory uncertainty describes the apparent variability and randomness of real-world systems (Bedford and Cooke 2001; Beer 2006; Skinner et al. 2014; Refsgaard et al. 2013; Beven 2016; Mishra, Karmakar, and Kumar 2018). Other labels have also been applied to this category including inherent, physical, or natural variability (Vesely and Rasmuson 1984; Finkel 1990; Baecher and Christian 2000; Walker et al. 2003); inherent randomness (Morgan and Henrion 1990); stochastic uncertainty (Helton 1994; Refsgaard et al. 2007); ontic uncertainty (Knol et al. 2009); and intrinsic uncertainty (Briggs, Sabel, and Lee 2009). Meanwhile, epistemic uncertainty describes uncertainty that arises due to knowledge deficits, which has also been called knowledge-based uncertainty (van Asselt and Rotmans 2002; Petersen et al. 2003; Ascough II et al. 2008), incertitude (Hayes et al. 2006), and extrinsic uncertainty (Briggs, Sabel, and Lee 2009).

While aleatory uncertainty is irreducible, that does not mean it is completely beyond our control. Beven (2016) describes aleatory uncertainty as uncertainty with “stationary characteristics,” meaning that it can often be described with a random distribution. Therefore, aleatory uncertainty can sometimes be managed or contained by representing it in a model as a random variable.

The central distinction between aleatory and epistemic uncertainty is that epistemic uncertainty describes reducible uncertainty while aleatory uncertainty describes fundamentally irreducible uncertainty that is an inherent property of the system. The purpose of the nature dimension is to sort all uncertainties into these two distinct boxes. Upon closer inspection, this tidy bifurcation break down.

The aleatory/epistemic distinction hangs on an ontological claim that some aspects of physical and social systems are irreducibly uncertain on a fundamental level. This position is supported by developments in the field of quantum mechanics in the first half of the 20th century, which many see as having closed the book on the notion of a completely calculable and knowable universe. However, prior to Heisenberg's uncertainty principle, the deterministic view of the universe laid out by classical mechanics and captured by Laplace's metaphor of an all-knowing "demon" went largely unchallenged. Laplacian determinism is elegantly described by Popper who compares it to a motion-picture film:

> [T]he picture or still which is just being projected is the present. Those parts of the film which have already been shown constitute the past. And those which have not yet been shown constitute the future ... Though the spectator may not know the future, every future event, without exception, might in principle be known with certainty, exactly like the past, since it exists in the same sense in which the past exists (1982, pp. 5).

But discoveries in quantum mechanics revealed the fundamental limitations of the accuracy of measurement and brought about a new scientific paradigm characterized by stochasticity and unknowability (particularly at the subatomic level). While this paradigm has been widely embraced and is embedded within the distinction between aleatory and epistemic uncertainty, it continues to be challenged by some theoretical physicists who advocate a more agnostic perspective (Svozil 1996; Casti 1996). For instance, Casti points out that the Heisenberg uncertainty principle, which describes the inherent limitations of measuring the physical world, is actually a limitation of particular mathematical formulations of quantum theory, which may or may not reflect a fundamental limitation in the real world itself. So, while theorists of uncertainty like van Asselt (2000, pp. 85) claim that: "Variability is an attribute of reality," variability may, in fact, reflect but one *perspective* of the attributes of reality.

More importantly, the aleatory/epistemic distinction often frays when applied to real-world examples. For example, the seemingly random susceptible-host transmission rate of a pathogen (like the novel coronavirus) may in fact be the result of complex within-host dynamics that are discovered by new technologies or methods (Gog et al. 2015). The seemingly random flocking behaviour of bird species like starlings can be largely explained by individual birds applying three simple rules to a dynamic environment (Hildenbrandt et al. 2010). While there may exist truly irreducible uncertainty, quite often, today's aleatory uncertainties turn out to be tomorrow's epistemic uncertainties.

Proponents of aleatory and epistemic uncertainty acknowledge that the distinction between the two is not always clear. Walker et al. admit that "it may be difficult to identify precisely what is reducible through investigations and research, and what is irreducible because it is an inherent property of the phenomena of concern" (2003, pp. 14). Similarly, Briggs et al. remark: "beyond the quantum scale, true randomness is surprisingly uncommon ... though this is often masked by the inherent complexities or lack of available data" (2009, pp. 191).

These arguments serve to challenge the immutability of the distinction between fundamentally reducible and irreducible uncertainties. A more defensible position might be to differentiate between reducible epistemic uncertainty on the one hand and uncertainty that *appears* to be random or fundamentally irreducible on the other. Formulated this way, the aleatory/epistemic distinction can still serve as a useful roadmap for modelers and decision makers by helping them to distinguish uncertainties that can be reduced from those that (probably) cannot.

However, this reformulation also assumes that all epistemic uncertainties are reducible—an idea that is challenged by chaotic deterministic systems and "NP-complete" problems. Chaotic deterministic systems lack closed-form solutions and their behaviour—which may be theoretically predictable (i.e. we know what steps must be taken to predict it)—may be impossible to predict within human-relevant timeframes (Crutchfield et al. 1986). Similarly, NP-complete problems[13] cannot be solved in polynomial time. For instance, it is estimated that the famous "travelling salesperson problem" would require more time to solve than the age of the universe, even with the fastest computers (Casti 1996). These examples introduce an entirely new category of uncertainty typically ignored by existing frameworks: *irreducible epistemic uncertainty*.

To summarize, the distinction between aleatory and epistemic uncertainty is, itself, uncertain (we lack confidence in the validity of this distinction). And therefore, the claim that aleatory and epistemic uncertainty constitute qualitatively distinct kinds of uncertainty is up for debate. However, the more important question is whether this distinction provides extra analytical value for modelers and decision makers. The practical value of distinguishing between aleatory and epistemic uncertainty appears to be that it helps modelers and decision makers differentiate between uncertainties that are worth trying to reduce and uncertainties that are not. As I argue in Section 2.5.3, this distinction is indeed quite useful—even though it is difficult to measure reducibility *ex ante*. However, rather than trying to tie reducibility to the intrinsic nature of different types of uncertainty, I believe reducibility fits more comfortably within the dimension of uncertainty level.

But this begs the question: should a conceptualization of uncertainty abandon attempts to distinguish between fundamentally different *kinds* of uncertainty altogether? Kwakkel et al.'s decision to add

[13] NP stands for "nondeterministic polynomial time." An NP-complete is an intractable problem where no efficient solution algorithm has been found.

ambiguity as a third nature of uncertainty in the updated W&H framework provides a promising alternative direction.

2.6.2 Ambiguity and linguistic uncertainty

Recent scholarship has emphasized two additional dimensions of uncertainty: linguistic uncertainty and ambiguity (which Kwakkel et al. group under the term ambiguity). Linguistic uncertainty stems from the inherent limitations of human language—that language is vague, imprecise, constantly evolving, and context-dependent (Ascough II et al. 2008; Regan et al. 2002; Elith et al. 2002; Beven 2016; Kujala et al. 2013; Morgan and Henrion 1990). Ambiguity refers to uncertainty that emerges from differences in beliefs, values, worldviews, and "the simultaneous presence of multiple valid and sometimes conflicting ways of framing a problem" (Brugnach and Ingram 2012, pp. 61. See also: Brugnach et al. 2008; Dewulf 2013; Dewulf et al. 2005; Finkel 1990; Refsgaard et al. 2013).[14]

Both linguistic uncertainty and ambiguity arise from the messiness and fuzziness of human interactions. These uncertainties tend to emerge from the modeling process itself, including the communication of model outputs, and the processes through which model outputs inform decision-making. Attempts to define components of a system are impeded by the fact that language can be like an ill-fitting garment: at once too stiff and constricting—like a glove with only four fingers—and too loose and underspecified—like a shirt several sizes too large. Language is a rather clumsy and imprecise technology for describing the detail and complexity of reality (Elith et al. 2002). Meanwhile, ambiguity can impede decisions about where to draw the boundaries of the system, as individuals with different knowledge frames, worldviews, or values may disagree about which aspects of the system should be the focus of attention (Brugnach et al. 2008; Rittel and Webber 1973).

Various uncertainty frameworks suggest that these uncertainties arising from human interactions possess properties that make them different *in kind* from other uncertainties. Both linguistic uncertainty (Ascough II et al. 2008) and ambiguity (Brugnach and Ingram 2012; Refsgaard et al. 2013) have been positioned as a third nature of uncertainty alongside epistemic and aleatory uncertainty. One core property of both linguistic uncertainty and ambiguity emphasized in these typologies is the difficulty or impossibility of reducing them. For linguistic uncertainty, this argument closely resembles the discussion of aleatory uncertainty. Efforts to reduce linguistic uncertainty (e.g. defining key terms) have decreasing

[14] The term ambiguity has also been used in a different sense to describe uncertainties with poorly defined probabilities (Beer 2006; Dequech 2011).

marginal returns. Like a curve approaching an asymptote, perfect confidence may be unobtainable (Fig. 4).

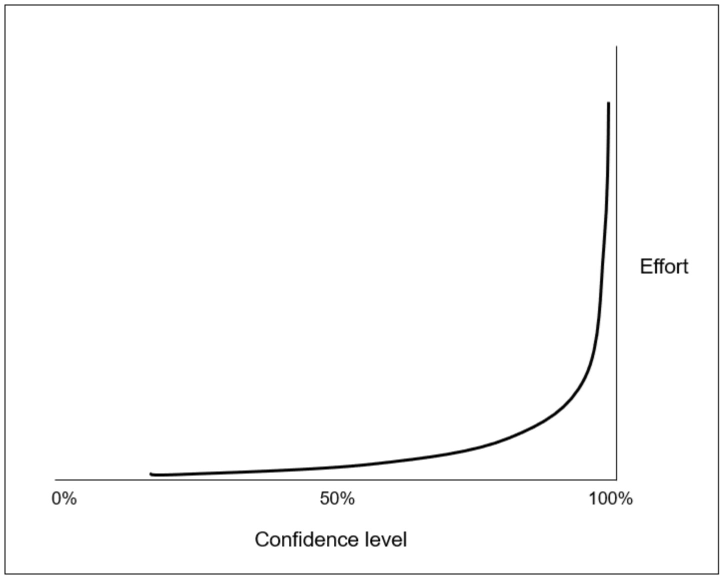

Figure 4: Fundamental limits to the reducibility of linguistic uncertainty?

As I argue in Section 2.5.3, declarations that some uncertainties are fundamentally irreducible are more usefully framed as being practically irreducible. With recent advancements in the development of neural networks and brain-to-brain interfaces, it is not too difficult to imagine a world where we are capable of significantly reducing linguistic uncertainty. However, such uncertainty reductions are unlikely to materialize in the near-future, and thus some amount of practically irreducible uncertainty is unavoidable when humans communicate with each other today.

Ambiguity is also difficult—or perhaps impossible—to reduce completely. Uncertainty reduction is concerned about increasing one's confidence in the "truthfulness" of one's understanding of reality. However, ambiguity arises from the fact that there may exist multiple equally valid but conflicting ideas about what is "true" when multiple agents are presented with the same information. While some uncertainties arising from the clash of perspectives or worldviews may be reducible insofar as misunderstandings can be identified and resolved, some differences—such as fundamental ontological and epistemological disagreements—may be essentially irresolvable.

For example, in Section 2.5.1 I make the case for interpreting uncertainty from a moderate subjectivist Bayesian perspective. However, many frequentists (or objectivist Bayesians) may disagree with the claim that likelihoods emerge from our minds and would assert that there are "correct" likelihoods attached to every phenomenon (even if we are limited in our ability to ascertain them). These differences between knowledge frames may not be practically reducible because, from our current vantage point, there does not appear to be a bridge for us to cross.

Of course, most people do not believe that all ontological positions, knowledge systems, or values are “equally valid.” Since knowledge production is an intersubjective process (from a Bayesian perspective), it is at least theoretically possible that in the distant future, humans could converge around a small number of knowledge frames (which is already the case in many epistemic communities), thereby reducing or eliminating ambiguity in many situations. But in many cases, ambiguity casts an omnipresent shadow over all attempts to define or model a system, identify possible outcomes, determine their relative likelihood, and communicate the significance of that information to others.

Putting the issue of fundamental limits to reducibility to the side, the property that best supports the claim that linguistic uncertainty and ambiguity together constitute a qualitatively distinct kind of uncertainty is that they both involve the uniquely reflexive nature of human systems. Reflexivity describes how circular or bidirectional causal relationships embed human beings within the systems they try to understand (Beinhocker 2013). Linguistic uncertainty and ambiguity both have an “inescapable” quality to them—all attempts to describe and reduce them necessarily involve the use of language and are informed by, and imbued with, a particular worldview or knowledge frame. Like trying to climb up a Penrose staircase (Fig. 5), no matter how much energy we expend attempting to pinpoint, measure, and reconcile uncertainties emerging from language and divergent knowledge frames, any perceived progress is, to some extent, illusory.

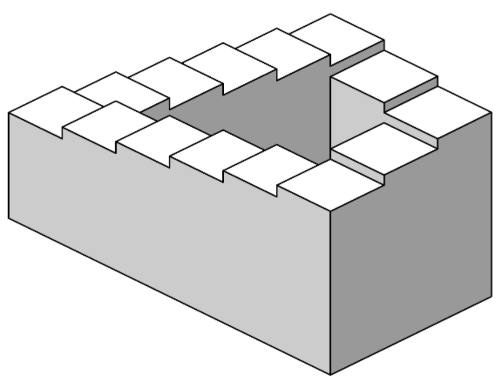

Figure 5: Penrose staircase (source: Wikimedia Commons)

Therefore, I agree with recent frameworks that ambiguity and linguistic uncertainty collectively qualify as a distinct and useful nature of uncertainty. However, if ambiguity and linguistic uncertainty emerge from the unique properties of human systems, then it stands to reason that uncertainties emerging from systems that lack the messiness produced by human interactions constitute a second *kind* of uncertainty. Recent contributions to the field of complexity science support the claim that there are fundamental differences between systems involving human agents and systems that do not—a claim that offers a novel and useful way to classify uncertainties into distinct natures.

2.6.3 Complex reflexive uncertainty

The field of complexity science is centrally concerned about what it means to say that one system is more complex than another. One particularly useful contribution to this conversation is made by Beinhocker (2013) who identifies distinct categories of systems and arranges them along a "spectrum of complexity" (Fig. 6). Beinhocker starts by making the basic distinction between simple mechanical systems that exhibit linear causation and complex mechanical systems that have non-linear dynamics and emergent behaviour. The implication is that the properties of non-linear dynamics and emergent behaviour are evidence of greater system complexity.

Complex adaptive systems (CAS) constitute an even higher rung on the complexity ladder. In addition to exhibiting non-linear causation and emergent behaviour, CAS can be distinguished from complex mechanical systems by the presence of interacting agents possessing internal models (sometimes called schemas). A CAS is a system that is "far from equilibrium, self-organising, and … co-evolves from the interaction between heterogeneous agents" (Fuller and Warren 2006, pp. 957). According to Beinhocker (2013, pp. 331), CAS possess four key characteristics:

- at least one *agent* interacting with an environment;
- a *cognitive function* that allows agents to receive information about the environment;
- a *manipulative function* that allows agents to interact with or change the environment; and
- a dynamic *internal model* or schema that connects agents' cognitive and manipulative functions.

This internal model is what allows agents to adapt their strategies and behaviours to changing environments. For example, seemingly simple organisms like bacteria are capable of coordinating the expression of different genes, thereby adapting to changes in the environment (Cunha, Xavier, and de Castro 2018).

Expanding on the work of Soros (2013), Beinhocker describes a subset of CAS called complex reflexive systems (CRS) that possesses even greater complexity than CAS. The purpose of this additional category is to account for the unique nature of human systems. A simplified version of Beinhocker's spectrum of complexity is presented in Figure 6 and the main characteristics of each system type are summarized in Table 3.

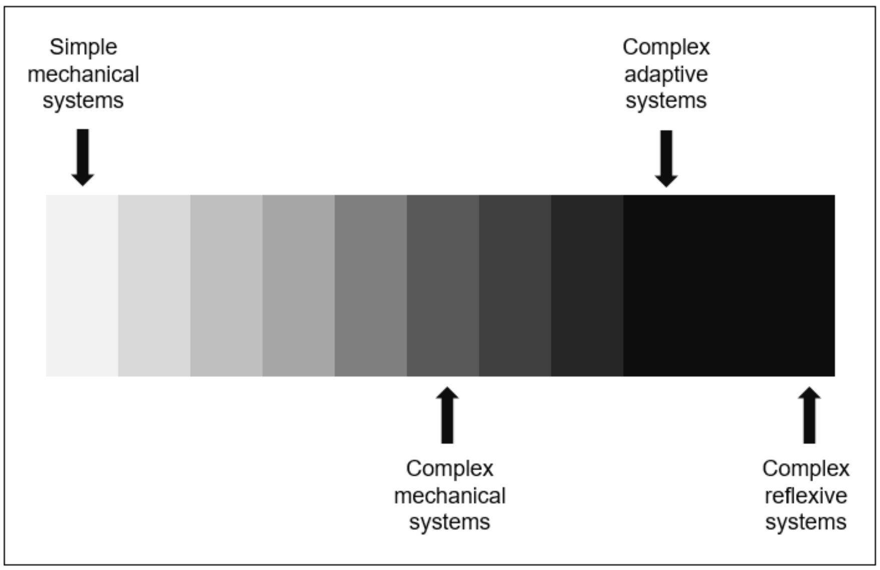

Figure 6: Spectrum of system complexity (adapted from Beinhocker 2013)

Table 3: Key characteristics of four levels of system complexity (adapted from Beinhocker 2013)

Complexity level	Characteristics	System example
Simple mechanical	• linear dynamics (proportional causation)	Internal combustion engine
Complex mechanical	• non-linear dynamics (feedback effects, disproportional causation)	Frictionless billiard ball table
Complex adaptive	• non-linear dynamics (feedback effects, disproportional causation) • interacting agents with internal models (schemas) • *internal models (schemas) adapt slowly* • *low interactive complexity*	Bacteria colony
Complex reflexive	• non-linear dynamics (feedback effects, disproportional causation) • interacting agents with internal models (schemas) • *internal models (schemas) adapt quickly* • *high interactive complexity*	National election campaign

Like CAS, CRS have agents with both cognitive and manipulative functions, coordinated by dynamic internal models. However, CRS differ from CAS in three ways. First, CRS exhibit much greater interactive complexity than CAS. Interactive complexity describes the density of interactions between heterogeneous agents within the system. While a bacteria colony can interact and coordinate by slowly altering their genes and collectively steering the colony away from threats, humans have myriad ways of interacting and coordinating within their environment including gene manipulation (which has greatly accelerated with recent advancements in gene editing technology) but also complex language, storytelling, and art. However, language is a highly effective but imperfect technology. The development of concepts and grammatical rules are attempts to corral and describe aspects of our perceived reality and are thus, like mathematical models, abstractions of an underlying reality—hence the emergence of linguistic uncertainty.

The second distinguishing characteristic of CRS is that their internal models of agents adapt at a much faster rate than CAS. This characteristic is connected to the high interactive complexity of CRS. Dense interactions between agents provide more opportunities for communicating information about the environment and strategies to manipulate it, thus facilitating the rapid coevolution of internal models. For example, the environmental movement that first emerged in the 1960s proposed a fundamental reframing of the relationship between human (economic and social) systems and natural systems. This radical shift in worldview (placing human social and economic systems within a broader "Earth system") has spread rapidly and may have the potential to produce significant behavioural change (Biermann 2012).

Third, arising from both the dense interactions of agents in CRS and the rapid evolution of internal models are many of the traits that we consider uniquely human, such as consciousness and the capacity for recursive thinking. Recursive thinking is what allows humans (and other complex mammals) to engage in "mental time travel" by constructing mental models of the past and future and drawing on those models to inform present decisions (Corballis 2011). It is this capacity that allows humans to construct value systems, ponder the nature of existence, and develop abstract concepts like knowledge, uncertainty, and truth. Ambiguity emerges from these unique traits of CRS. Without high interactive complexity, we would lack the linguistic sophistication necessary to articulate intricate worldviews and perspectives. Without the rapid evolution of internal models, communities of agents would not produce distinct knowledge frames. And without the ability to engage in mental time travel, we would be unable to ponder the mysteries of the universe and develop a plurality of ontologies and belief systems.

These key differences between CRS and non-reflexive systems (CAS and mechanical systems) provide a framework for classifying two distinct *kinds* or natures of uncertainty. Since both ambiguity and linguistic uncertainty emerge from the distinct characteristics of CRS, we can distinguish between uncertainty that emerges from CRS—what I call *complex reflexive uncertainty*— and uncertainty that emerges from non-reflexive systems—what I call *non-reflexive uncertainty*. This distinction provides a convenient way to group ambiguity and linguistic uncertainty together as one nature of uncertainty. And importantly, it also defines what makes this kind of uncertainty so difficult to reduce: the inherent messiness of human interactions.

The qualitative distinction between complex reflexive and non-reflexive uncertainty is useful for modelers and decision makers for two reasons. First, acknowledging that a particular uncertainty emerges from a CRS, a modeler or decision maker should be alerted about the unique reducibility challenges associated with ambiguity. As opposed to other uncertainties that can be reduced by making investments of effort that result in confidence gains, ambiguity is about normative disagreements between individuals or groups, or other potentially intractable disagreements about the nature of reality and knowledge. Therefore, ambiguities from a CRS may require specific strategies to reveal and resolve normative and ontological conflicts—or they may lie entirely beyond the pale of uncertainty reduction.

Second, the distinction between complex reflexive and non-reflexive uncertainty should alert modelers and decision makers to some of the practical challenges associated with trying to measure complex reflexive uncertainty. When it comes to anticipating the behaviour of a CRS, the combinatorial potential created by the rapid adaptation of schemas possessed by human agents makes it extremely difficult for a modeler or decision maker to have high confidence in the model structure or that all possible outcomes have been identified, leading to states of deep or Knightian uncertainty. Further, since there are currently tremendous limitations in our ability to accurately model many important psychological and cognitive dimensions of human systems such as ideology and identity (Homer-Dixon et al. 2013), the task of reducing this kind of uncertainty is often extremely effortful, requiring theories, methods, and data that we do not currently possess. In other words, acknowledging that uncertainty is emerging from a CRS can help modelers and decision-makers make better assessments of the level of uncertainty.

Here, I present complex reflexive uncertainty and non-reflexive uncertainty as distinct natures of uncertainty but the boundary separating CRS and CAS is somewhat porous. While complex reflexive systems are typically associated with human agents, it is possible that systems without human agents like advanced mammals (e.g. porpoises and primates), artificial intelligences, and even ecosystems may exhibit the interactive complexity associated with complex reflexive systems (Beinhocker 2013).

Lastly, it should be stated that every system that is modelled (even those that do not have a human dimension) is, in some way, "touched" by reflexive processes—simply by being observed and interpreted by a human agent. Therefore, all modeled systems—even simple mechanical systems—may be affected by linguistic uncertainty and ambiguity.

2.7 An amended uncertainty framework

The amended uncertainty framework presented in Figure 7 proposes that all uncertainties can be evaluated using the three dimensions of location, level, and nature.

First, uncertainties can be said to arise from one or more of the six locations identified by Kwakkel et al (2010): system boundary; conceptual model; computer model (structure and parameters); input data; model implementation, and; processed output data.

Second, uncertainties can be measured using two scales that describe how uncertain they are. The level dimension is made up of two distinct sub-dimensions: predictability and reducibility. Predictability can be evaluated on a four-level scale (shallow, medium, deep, and Knightian) describing the situation's uncertainty structure. The level of medium uncertainty is deemphasized in the figure to reflect that it is more accurately described as a sublevel of shallow uncertainty. Meanwhile, uncertainty reducibility can be classified as either practically reducible or practically irreducible depending on a decision maker's assessment of the evidence required to produce a meaningful increase in their confidence and the amount of effort required to produce that evidence.

Lastly, all uncertainties can be said to possess one of two natures: complex reflexive uncertainty or non-reflexive uncertainty. Linguistic uncertainty and ambiguity are positioned as sub-types of complex reflexive uncertainty.

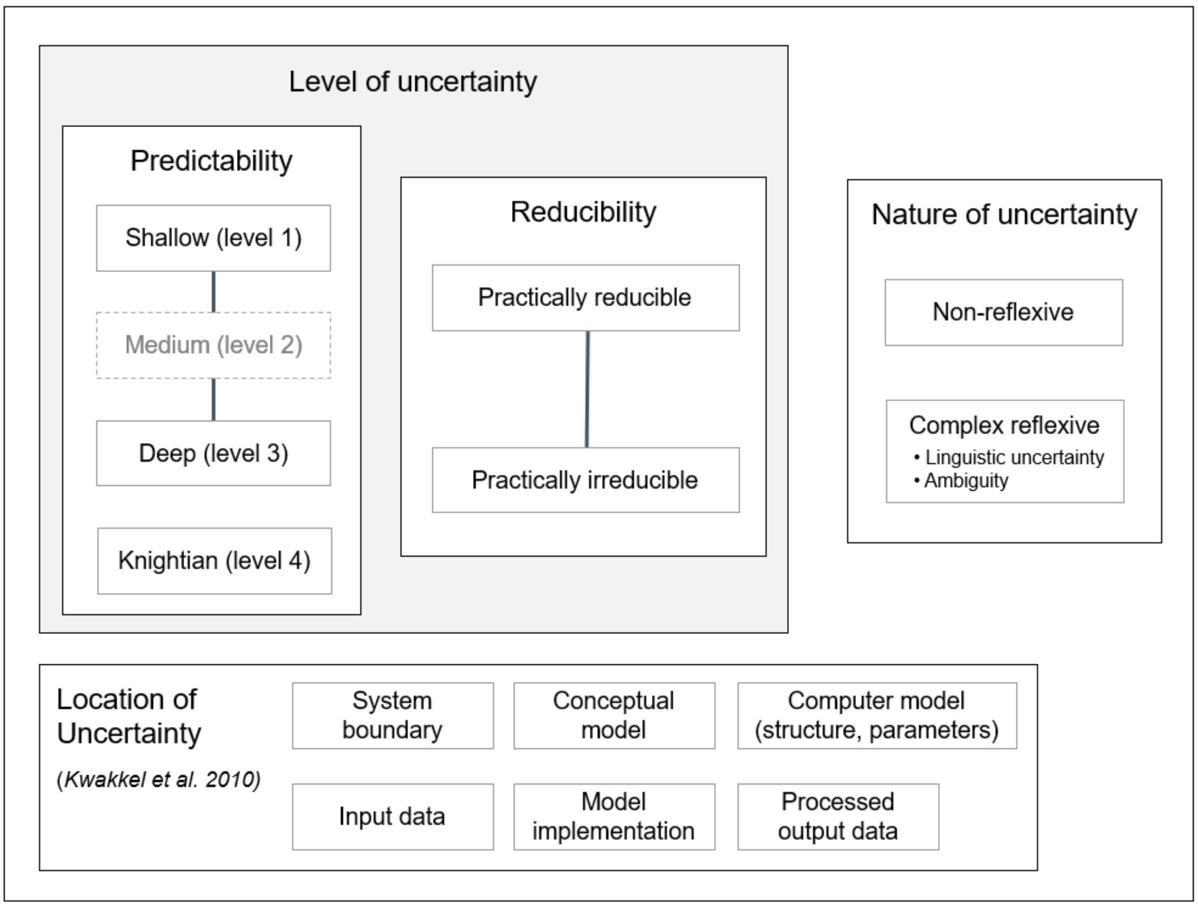

Figure 7: Amended uncertainty framework

2.8 Conclusion: implications for decision-making under conditions of uncertainty

This paper addresses the epistemological blind spot in the conceptual literature on uncertainty. Existing frameworks either avoid specifying criteria for when it is appropriate to express uncertainty probabilistically or perform a sleight-of-hand and flip from a frequentist perspective to a Bayesian perspective when frequentist methods become untenable. As I argue in Section 2.5.1, a consistent epistemological perspective increases the clarity and transparency of the framework without sacrificing its utility (due to the "epistemological agility" of the Bayesian perspective, whereby frequencies can be incorporated into subjective judgements). The commitment to a Bayesian perspective also leads to more fruitful discussions of concepts like reducibility and ignorance. Uncertainty reductions can be usefully reframed as investments of energy that result in increases in confidence. Meanwhile, understanding uncertainty as a mental state of imperfect confidence leads to the conclusion that "true" ignorance (which lies beyond our awareness) is not a dimension of uncertainty at all.

However, the real test for a conceptual framework of uncertainty is whether it is useful and adds analytical value that alternative frameworks do not. The amendments proposed in this paper point to a number of implications for decision makers navigating uncertainty. Despite the fact that measurements of uncertainty (i.e. likelihood assessments) are inherently subjective or intersubjective, they can be tethered to three different uncertainty structures: shallow uncertainty, deep uncertainty, and Knightian uncertainty. The main implication of this conceptualization of uncertainty level is that measuring the predictability of an uncertainty situation is a two-step process. When conducting likelihood assessments of different possible outcomes, decision makers first assess their confidence level that all possible outcomes have been identified. Only if they have high confidence that all possible outcomes have been identified (which could mean establishing upper and lower bounds) should they then proceed to assess their confidence level in the relative likelihood of different outcomes and select a measurement scale that aligns with their confidence level. This two-step process is particularly important for systems involving human agents, since the combinatorial possibilities of complex reflexive systems can be staggering and it is more likely that the decision maker is ignorant of possible outcomes.

Another important implication of the conceptualization of uncertainty presented in this paper is that decision makers should be aware of the deceivingly complex and challenging process of estimating the reducibility of uncertainty situations *ex ante*. Estimating uncertainty reducibility is a multi-step modeling process where the decision maker first constructs a model describing the evidence that they believe will lead to confidence increases in the extent or quality of decision-relevant knowledge. This model is then incorporated into an evidence-effort model which describes the decision maker's estimates of the effort necessary to gather the evidence. However, the decision maker may lack confidence in the accuracy of either of these models and these confidence deficits may, in turn, require their own uncertainty reduction activities (and models).

Finally, further attention should be paid to how these evidence and evidence-effort models underlie important R&D priority-setting decisions across a broad range of policy domains. One of the main ways that organizations and individuals grapple with uncertainty reduction is through the process of prioritizing uncertainty reduction activities (i.e. R&D projects). R&D priority-setting describes the process of allocating finite resources to systematically reduce uncertainties that are impeding strategic decision-making. Determining which uncertainty reduction activities should be funded and which ones should not requires complex cost-benefit calculations and *ex ante* estimates of the relative importance and reducibility of different uncertainties that are impeding decision-making.

Chapter 3

Article #2: R&D priority-setting for global catastrophic risks: the case of the NASA planetary defense mission

3.1 Preface

Decision-making around what to do about global catastrophic risks (GCRs) is impeded by the presence of confidence deficits (i.e. uncertainty). Experts and policy makers are uncertain about the nature of the problems they are facing, the effectiveness and feasibility of possible solutions, and how to overcome technological, economic, and political obstacles to implement those solutions. Fortunately, uncertainty can often be reduced (i.e. confidence in one's knowledge can be increased) by learning. Organizations learn by conducting research and development (R&D). R&D projects are effortful – they require money, time, and human capital. But they also have the potential of producing better decisions, and thus could create social value. These cost and benefit estimations are at the heart of R&D priority-setting – a process that every organization conducting research with a limited budget must execute. In this article, I approach R&D priority-setting as a process of systematic uncertainty reduction.

This article explores the process of systematic uncertainty reduction around planetary defense (i.e. near-Earth object impact prevention) at the National Aeronautics and Space Administration (NASA). It is, first and foremost, a descriptive case study. It describes how planetary defense R&D decisions are made at NASA and what the outcomes from those decisions looked like in 2019. A version of the article will be published in the journal *Research Policy* in July 2021 (Janzwood 2021).

In the context of the larger dissertation, this article tackles one of the two main decision support tasks: reducing uncertainty (the other being the communication of uncertainty). A closer look at the decision analysis literature reveals that there are many sub-tasks involved in the systematic reduction of uncertainty, including: defining the problem, identifying uncertainties and candidate projects, estimating benefits and costs, and selecting and executing uncertainty-reducing projects. The exploration of the uncertainty concept in Article #1 provides an important conceptual foundation for many of these sub-tasks. Article #1 also describes how the estimation of benefits involves the construction of "evidence models" describing the types of evidence that would significantly increase one's confidence that one's decision-making would be improved. "Evidence-effort models" are also required to describe how costly that evidence would be to obtain. This perspective guides my investigation of the processes that research managers at NASA follow to evaluate and select the R&D projects that they hope will improve decision-making around planetary defense the most (within their budgetary constraints).

A significant amount of research for this article was undertaken at the 2019 International Academy of Astronautics (IAA) Planetary Defense Conference where I conducted a survey collecting expert opinions on planetary defense R&D priorities. I am indebted to the individuals that took the time to take the survey, as well as the senior managers and directors at NASA that participated in interviews.

3.2 Abstract

"Mission-oriented" public research organizations invest in R&D to improve decision-making around complex policy problems from climate change to asteroid impacts, thus producing "public value." However, the estimation of benefits produced by such R&D projects is notoriously difficult to predict and measure – a challenge that is magnified for GCRs. GCRs are highly uncertain risks that may pose enormous negative consequences for humanity. This article explores how public research organizations systematically reduce key uncertainties associated with GCRs. Building off of recent literature highlighting the organizational and political factors that influence R&D priority-setting at public research organizations, this article develops an analytical framework for explaining R&D priority-setting outcomes that integrates the key stages of decision analysis with organizational and political dynamics identified in the literature. This framework is then illustrated with a case study of the NASA planetary defense mission, which addresses the GCR of near-Earth object (asteroid and comet) impacts. The case study reveals how organizational and political factors interact with every stage in the R&D priority-setting process – from initial problem definition to project selection. Lastly, the article discusses the extent to which the case study can inform R&D priority-setting at other mission-oriented research organizations, particularly those addressing other GCRs.

3.3 Introduction

Public organizations invest in research and development (R&D) to address complex societal challenges where decision-making is impeded by uncertainty. Investment in R&D serves to reduce uncertainty, thereby improving decision-making and producing "public value." Recent scholarship on organizational priority-setting and budget allocation has highlighted the unique challenges facing public research organizations compared to corporations, particularly the challenge of quantifying public value and incorporating it into cost-benefit analysis (CBA) (Ciarli and Ràfols, 2019; D'Este et al., 2018; Wallace and Rafols, 2015, 2018). The challenge of measuring the public value of R&D is magnified for organizations addressing a subset of policy problems known as global catastrophic risks (GCRs).

GCRs are risks of events that could significantly harm or destroy human civilization on a global scale (Hempsell 2004a; Baum 2010; Barrett 2017). GCRs that pose a potential threat to humanity include pandemics, near-Earth object (NEO) impacts, climate change, nuclear war, bioterrorist attacks, and artificial intelligence (AI) accidents. R&D investments addressing GCRs could help significantly improve decision-making related to preventing catastrophic events or decreasing their impact and thus could yield enormous public benefits. This article responds to calls for improving the state of knowledge on what governments and other organizations are currently doing to address GCRs (Currie and Ó hÉigeartaigh 2018; Dafoe 2018b) and argues that R&D priority-setting is an important and underappreciated form of GCR governance.

Much of the R&D priority-setting literature discusses the utility of decision analytic methods for estimating the performance of R&D investments and improving decision-making (e.g. Keisler 2004; Bates et al. 2016; Barrett 2017; Drago and Ruggeri 2019; Bhattacharjya, Eidsvik, and Mukerji 2013; Arratia et al. 2016). The decision analytic perspective provides insight into the many sub-tasks that all organizations perform when making R&D budget decisions such as defining the problem and performing cost and benefit estimations. Meanwhile, another strand of scholarship emphasizes how organizational and political factors also play a significant role in shaping research priorities, particularly for public organizations (Brattström and Hellström 2019; Ciarli and Ràfols 2019; Hellström, Jacob, and Sjöö 2017; Cruz-Castro and Sanz-Menéndez 2018; D'Este et al. 2018; Wallace and Ràfols 2018, 2015).

This article seeks to bridge the two perspectives by proposing an analytical framework for describing and explaining R&D priority-setting outcomes at "mission-oriented" research organizations, such as those addressing GCRs. The framework integrates the key stages of decision analysis with organizational and political factors. Instead of viewing organizational and political factors as a filter that R&D priorities pass through at the very end of the priority-setting process, the framework proposes that they are present at every stage—from initial problem definition to final project selection.

This analytical framework is then demonstrated with a case study of NASA's planetary defense mission, which aims to reduce the risk posed by NEO (asteroid and comet) impacts with the Earth. The study responds to a call for more descriptive research on the use of decision analysis in organizational settings (Kleinmuntz, 2007) and builds on recent studies examining the priority-setting role of program managers at the agency level (Brattström and Hellström 2019; Hellström, Jacob, and Sjöö 2017; Wallace and Ràfols 2018).

Through a multi-method research design that combines bibliometric analysis, survey data analysis, and semi-structured interviews with senior research managers at NASA, the case study describes the process through which planetary defense R&D project proposals are evaluated and prioritized. Budget allocation decisions for planetary defense at NASA emerge from an iterative process of peer review and expert consultations, with final authority typically resting with a small number of senior NASA administrators. Senior research managers attribute discrepancies between the research priorities articulated by the 2019 NASA planetary defense budget and the research priorities of the broader planetary defense expert community to a number of factors, including an institutional bias within NASA that privileges basic science over more applied R&D. R&D decisions are occasionally dictated by the White House or Congress, bypassing the standard prioritization process entirely. Therefore, this case study illustrates how R&D priority-setting decisions at mission-oriented organizations emerge from continuous interactions between formal priority-setting processes, organizational factors, and exogenous forces outside of the organization. It also provides insights into the challenges associated with R&D priority-setting that may be useful for organizations addressing other GCRs.

Section 3.4 describes the unique R&D priority-setting challenges presented by GCRs and frames R&D priority-setting as a process of strategic uncertainty reduction. Section 3.5 proposes an analytical framework for describing and explaining R&D priority-setting outcomes for public organizations, while Section 3.6 uses this framework to explain R&D priority-setting outcomes at NASA around planetary defense. Lastly, Section 3.7 discusses the extent to which the case study can inform R&D priority-setting at other mission-oriented organizations, particularly those addressing GCRs.

3.4 Strategic uncertainty reduction for global catastrophic risks

R&D funding decisions at mission-oriented organizations can be understood as strategic investments in uncertainty reduction. The benefits created by R&D investments stem from the reduction of various uncertainties and the resulting improvements in decision-making. In Article #1, I define uncertainty as *a mental state of imperfect confidence in the extent or quality of one's knowledge*. If uncertainty is a state where we lack perfect confidence in the extent or quality of our knowledge, then a reduction of uncertainty is a step towards perfect confidence.

For organizations addressing GCRs, these confidence gains are usually not estimated by a single individual. Typically, R&D funding decisions – and, by extension, estimations of potential confidence gains – are influenced by multiple actors within the organization. For instance, a committee voting on the relative costs and benefits associated with various R&D projects produces an intersubjective assessment of the extent to which uncertainty will be reduced and how valuable that reduction is. Typically, certain actors within the organization like research directors and senior leadership have more influence over these decisions than others.

GCR describes the class of risks facing humanity that could produce a significant decrease in global population or human welfare, the collapse of civilization as we know it, or human extinction. There is some controversy about where exactly to draw the threshold separating GCRs from "sub-catastrophic" risks (Avin et al. 2018; Baum and Handoh 2014). However, GCRs are generally understood as unprecedented risks, which suggests that the impacts of a GCR would exceed the estimated 35 to 60 million deaths (1.5-2.6% of global population) that occurred during World War II (Royde-Smith and Hughes 2019) or the 50 to 100 million deaths (2.5-5% of global population) attributed to the 1918-1921 influenza pandemic (Johnson and Mueller 2002). A list of GCRs is provided in Table 1.

GCRs are prime candidates for R&D investment because they tend to be linked to large portfolios of uncertainties that hinder decision-making. R&D that is specifically directed towards helping solve complex policy problems is sometimes referred to as "mission-oriented R&D" (Mowery 2012; Wallace and Ràfols 2015; Robinson and Mazzucato 2019). A distinction is commonly made by research organizations like the US National Science Foundation between basic research, applied research, and development (Kennedy 2012), with basic research being closer to what is commonly labeled "science," development referring to the creation of new technologies, and applied research occupying a grey middle-area (Avin 2014). Mission-oriented research is typically understood as encapsulating technology development and applied research. However, the category of basic research is sometimes expanded to include "strategic science," which is basic research undertaken with the expectation that it will likely support solutions to practical problems (Irvine and Martin 1984). Therefore, mission-oriented R&D can entail (strategic) basic research, applied research, and technology development.

Of course, the presence of uncertainties impeding decision-making is not unique to GCRs. While some policy problems that pose only sub-catastrophic impacts also meet the following criteria, GCRs tend to possess four characteristics that amplify the importance of R&D and R&D priority-setting:

1. the timeline for a risk event is believed to be non-imminent or uncertain;

2. key decisions are impeded by the presence of uncertainty;
3. uncertainty-reducing R&D projects are expensive or time-consuming; and
4. the benefits produced by uncertainty-reducing R&D projects are, themselves, deeply uncertain.

First, key decisions about how to avoid or decrease the impacts from a risk event must be relatively non-urgent or highly uncertain for investments in uncertainty reduction to be valuable. If a catastrophic asteroid impact is expected to occur tomorrow, it hardly makes sense to invest in R&D projects that would help scientists better understand the composition and orbital dynamics of NEOs. Even for climate change, where devastating impacts are already manifesting, truly catastrophic impacts are not expected for several decades even under the most pessimistic scenarios (Steffen et al. 2018). The general belief is that the window of opportunity is still open (but closing rapidly) to invest in uncertainty reductions that could help avoid the worst climate impacts. That is not to say that all GCRs are *necessarily* non-imminent. But currently, GCRs tend to have perceived time lags between when scientists first became aware of the risk and when they believe catastrophic impacts will manifest, increasing the value of strategic uncertainty reduction.

Second, R&D – and by extension R&D priority-setting – is only valuable if decisions about how to respond to a risk are impeded by various uncertainties. The more obstructive the uncertainties, the more valuable investments are to reduce or eliminate them. Decision-making around GCRs is certainly impeded by many forms of uncertainty. These uncertainties may also interact with and reinforce one another, further complicating decision-making.

Third, R&D priority-setting becomes increasingly important when resources to invest in socially valuable uncertainty reductions are scarce. GCR mitigation strategies – like transforming the global energy system to address climate change or developing viable NEO deflection technologies – tend to be expensive and difficult to produce. The fewer resources there are to invest in R&D projects, the more important it is to prioritize the projects that will improve decision-making the most.

The first three characteristics are typical of many complex policy problems, whether they pose catastrophic impacts or not. However, the fourth characteristic – that the benefits produced by uncertainty-reducing R&D projects are, themselves, highly uncertain – is particularly acute for GCRs because they are capable of producing unimaginable harm. Consequently, the social value produced by avoiding such risks is tremendously difficult to calculate.

Barrett (2017) notes that estimates of the value of preventing global catastrophe vary wildly from as low as $10 billion (Bostrom and Ćirković 2008) to infinity (Weitzman 2008; Baum 2010). Standard cost-benefit analysis would suggest that humanity should be willing to pay an amount equivalent to the expected harm of a risk event. However, estimates of both the likelihood that GCRs will occur within the next century and how to calculate the harm from a significant decrease in global population are generally clouded by deep uncertainty. Therefore, the true value of a GCR reduction activity is extremely difficult to estimate and compare against its costs – particularly given the fact that there are plenty of sub-catastrophic policy problems with much clearer estimates of costs and benefits that are also competing for the same finite resources.

While the benefits produced by GCR reduction activities are uncertain, the benefits produced by a single uncertainty-reducing R&D project can be even more difficult to calculate. Key decisions (i.e. the decisions that have a direct effect on decreasing the likelihood or impact of a risk event) are usually impeded by multiple uncertainties—each of which contains a series of sub-uncertainties. At best, an individual R&D project can hope to reduce or eliminate one or two of these uncertainties—and therefore, benefit estimations of R&D projects are estimations of partial improvements in decision-making.

Take, for example, the simple decision of whether to grab an umbrella when leaving your home. This decision is impeded by uncertainty about whether there is a significant likelihood of rain—or if you even care about getting wet. Uncertainty about the likelihood of rain contains sub-uncertainties, such as uncertainty about which meteorologist produces the highest quality probability of precipitation (POP) estimate and what POP constitutes a "significant" chance of rain. Your assessment of the quality of POP estimates from different meteorologists may, in turn, be influenced by how accurate they have been in the past.

Therefore, an R&D project investigating the reliability of meteorologists in the past would contribute to reducing uncertainty about which POP estimate is best—but it would not help reduce uncertainty about what POP constitutes a significant chance of rain nor whether you even care about getting wet in the first place. The R&D project only partially improves your decision-making about whether or not to grab an umbrella. This chain reaction or "trickle-up" effect, where reductions of second- and third-order uncertainties contribute to reductions of first-order uncertainties, is illustrated in Figure 8.

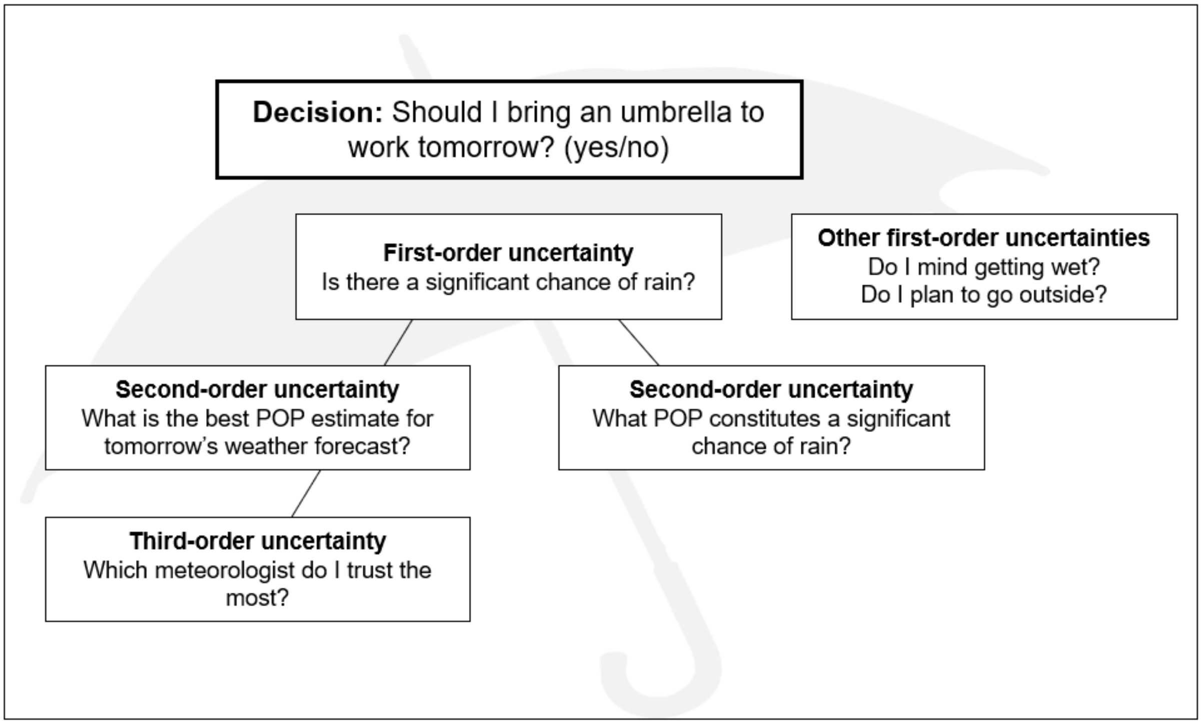

Figure 8: Uncertainty hierarchy example

The final point I want to make about R&D priority-setting as a process of systematic uncertainty reduction is about the use of the term "portfolio." "Uncertainty portfolios" and "R&D portfolios" are distinct but related concepts. Here, the term uncertainty portfolio refers to the collection of knowledge deficits impeding decision-making, while an R&D portfolio is widely used in the literature to refer to the collection of R&D projects being funded by an organization. A single research project may reduce multiple uncertainties at the same time, while the reduction of a single source of uncertainty may vastly reshape the benefits produced by a project in one's R&D portfolio.

Only three studies address the role played by R&D and R&D priority-setting in efforts to understand and address GCRs. Barrett and Baum (2017) and Barrett (2017) apply value of information approaches to AI risk and the NEO impact hazard respectfully, while Lee et al. (2014), conduct a six-step decision analysis of the NEO impact hazard. While all three studies strongly support the applicability of decision analytic methods to GCR research governance, none of them describe the current state of planetary defense R&D priority-setting, which is the central task of Section 3.6.

In order to explain R&D priority-setting outcomes at a mission-oriented organization, it is important to not only address the processes implemented by research managers to systematically measure and compare the value of R&D projects, but also the organizational and political factors that shape R&D decisions.

3.5 Explaining R&D priority-setting outcomes at mission-oriented organizations

This section examines the two dominant strands of scholarship addressing R&D priority-setting: (1) the decision analysis literature and (2) research in the field of science and technology (S&T) policy on organizational and political influences on priority-setting. I argue that both perspectives provide valuable insights for explaining R&D governance in mission-oriented organizations. I then present an analytical framework for explaining how R&D decisions are made at mission-oriented organizations that integrates political and organizational factors with the decision analysis cycle.

3.5.1 Decision analysis

At its core, the field of decision analysis is concerned with methods and processes that help to achieve "clarity of action," thereby increasing the value produced by decisions (Howard 2007). Clarity of action can be increased by making investments in R&D that reduce uncertainty. Thus, decision analytic methods are generally focused on estimating the value of information produced by R&D projects, which is often referred to as the expected value of perfect information (EVPI).

First introduced by Raiffa and Schlaifer (1961), the concept of EVPI describes the increase in expected value if all uncertainty is eliminated before making a decision (Yokota and Thompson 2004). It also represents the upper limit of how much should be spent on uncertainty reduction (Barrett and Baum 2017). Uncertainty reductions—which can be achieved through investments in R&D—are often expensive, time-consuming, and difficult. Therefore, EVPI estimations are conducted to determine whether R&D projects are worthwhile.

However, rigorous EVPI estimations can themselves be expensive, time-consuming and difficult. In fact, the cost of undergoing a formal EVPI analysis may exceed the benefits it produces (Keisler 2004). Therefore, the decision analysis literature describes a wide array of prioritization methods that organizations employ ranging from more informal, low-cost strategies like sorting or ranking alternatives to more disciplined and expensive EVPI approaches.

Lee, Jones, and Chapman (2014) identify a common set of tasks all research organizations execute when developing priorities, evaluating potential projects, and making budget decisions. While organizations will differ on the extent to which they apply rigorous analysis to each task, all of these tasks are necessary for systematically evaluating and prioritizing R&D projects. These tasks include:

1. defining the problem and objectives;
2. identifying decision options, uncertainties, and candidate projects;
3. modeling the problem (including the estimation of costs and benefits);
4. conducting sensitivity analysis to determine the impact of different uncertainties on estimated costs and benefits; and
5. selecting which projects to fund and how many resources to allocate to each.

3.5.1.1 Defining the problem, objectives, and major decision points

Characterizing the problem, articulating the organization's core objectives, and identifying the major decision points for dealing with the problem (which are often grouped together under the label "decision structuring" (von Winterfeldt and Edwards 2007)) are a crucial—but often overlooked (Neff 2014)—first step for priority-setting. The science and technology studies literature highlights how scientific knowledge does not exist in a vacuum but rather is embedded within a particular institutional, social, and cultural context (e.g. Jasanoff 2004). Consequently, problem or risk definition is also a fundamentally intersubjective process shaped by social factors (Renn 2008). Agreeing that a particular risk exists requires a degree of shared knowledge and common beliefs between members of a community. These include: a shared understanding of what the community values (e.g. continued existence, a base level of well-being, equality, freedom, etc.), the underlying systems that are responsible for producing or maintaining those values, and a causal story about exogeneous or endogenous dynamics that may disrupt or threaten those systems.[15]

These causal stories that connect a possible risk event to the negative impacts it could produce have both an empirical and a moral dimension (Stone 1989). For example, the number of casualties and injuries produced by a medium-sized NEO impact is an empirical question, but the value the international

[15] The identification of a collectively-perceived threat requires a common notion of: (1) the individuals that make up a community, (2) the shared interests or values of that community, (3) the physical and non-physical elements of the system (and their interactions) associated with increasing, maintaining or protecting shared interests or values, (4) the physical and non-physical elements of the system (and their interactions) that are responsible for threatening shared interests or values, and (5) a distinction between scenarios or possible worlds where the threat does, and does not exist. Taken together, these ingredients constitute an approximation of how the threat might play out in the "real world," including the key elements of the system and the rules governing how they interact.

community places on avoiding such a disaster (which might be confined to a single region) is a moral question. Therefore, for public research organizations, the task of defining a problem and the objectives for addressing it is a dynamic and social process involving a range of contextual factors both within and outside of the organization. The key decision points about how to respond to the risk emerge from both the framing of the problem and the organization's particular objectives.

3.5.1.2 Identifying uncertainties and candidate projects

Once key decision points have been defined by an organization, such as whether or not to initiate an NEO deflection mission, then uncertainties impeding those decisions—like the size, composition, and impact probability of the NEO—must be identified and described. As discussed in Section 3.4, uncertainties can be broken down into nearly endless levels of sub-uncertainties, which can be usefully classified into a hierarchy of "first-order" uncertainties, "second-order" uncertainties, etc. The complete list of uncertainties from various levels that collectively impede decision-making constitutes the organization's "uncertainty portfolio." An organization must then identify uncertainty reduction activities (R&D projects) that address its uncertainty portfolio. The identification of candidate projects can either be accomplished by people within the organization or by soliciting the help of outside experts. The remaining R&D priority-setting steps describe the subsequent analysis and management of an organization's candidate projects.

3.5.1.3 Estimating benefits

The benefits from R&D projects come from the reduction of various uncertainties impeding decision-making (i.e. increases in a decision maker's confidence in the quality and extent of their knowledge). Good decision-making produces social value and bad decision-making produces social costs. However, estimating the improvements in decision-making produced by a specific reduction in uncertainty is difficult. Complicating matters further, it is often uncertain whether an R&D project will deliver the expected uncertainty reductions claimed by its proponents.

Estimating the social benefits of R&D can be significantly more difficult to estimate than the anticipated returns from financial investments. The social value produced from uncertainty reductions is notorious for its lack of "convenient mathematical properties" (Hazen & Sounderpandian, 1999, pp. 126). Typically, EVPI is estimated using subjective utility functions (Barrett and Baum 2017; Edwards, Miles Jr., and von Winterfeldt 2007). However, as Claxton (2008) notes, EVPI is only useful if it can capture all relevant forms of social value produced by improved decision-making and takes into account all sources of uncertainty that may be distorting estimates of net social value. While it would be impossible for everyone affected by a GCR to agree on a valuation of the benefits produced by an improvement in decision-making around climate change or a disease outbreak, mission-oriented research organizations are charged with precisely this task. Without some estimation of the relative value of different R&D projects, organizations may as well allocate their resources randomly.

While sensitivity analysis is typically considered its own step in the decision analysis process, it ultimately contributes to refinements of benefit estimations. Sensitivity analysis assesses the effect of parameter uncertainty on decisions, which is analogous to the process of estimating “trickle-up” benefits discussed in Section 2. The trickle-up effect describes how R&D projects often reduce lower-order uncertainties, which then partially contribute to reductions of higher-order uncertainties that bear more directly on key decisions. With precise uncertainty values, decision makers can use Monte Carlo simulations to measure the relative importance of different uncertainty reductions on decision-making (Lee, Jones, and Chapman 2014). However, in cases that involve complex and imprecise measurements of social value, the trickle-up effect can only be roughly estimated.

3.5.1.4 Estimating costs

R&D projects that meaningfully reduce uncertainty are often expensive. With limited resources, organizations must make estimates of the amount of resources, effort, expertise, and time that are necessary to bring about the anticipated benefits of R&D projects. Typically, this process involves estimating the cost of labour, equipment, and external expertise, as well as the costs associated with monitoring, evaluation, and project risk. Costs associated with project risk are usually addressed by “risk adjusting” the valuation of projects, which can be done by applying a higher discount rate for riskier projects and a lower rate for less risky projects (Kleinmuntz 2007).

Martino (1995) describes two main costs associated with failing to select the best projects for an R&D portfolio. The first cost is the resources wasted on poor projects that yield negligible improvements to decisions. The second, even greater cost is the opportunity cost associated with better projects that were misevaluated or overlooked. With the low but non-negligible probability of a GCR manifesting in the near-future, the opportunity cost of missing out on a high-value project could be extremely high. In fact, a mission-oriented research organization’s ability to select an R&D portfolio with the absolute highest social value could be the difference between effectively responding to a GCR and experiencing a catastrophic event.

3.5.1.5 Integrating costs and benefits, selecting projects, and budgeting

The final step of the priority-setting process is integrating cost and benefit estimations into a CBA, selecting the optimal R&D portfolio, and allocating resources to each project. Here, dependencies and interaction effects between projects—both in terms of their costs and benefits—must be factored in (Bhattacharjya, Eidsvik, and Mukerji 2013; Kavadias and Loch 2004). The total social value produced by the ten projects that individually have the highest expected value per cost may be less than a different combination of ten projects where strategic overlaps produce economies of scale or benefit amplification effects. Timing is also an important factor for project selection (Barrett 2017; Martino 1995). An R&D portfolio that underperforms other portfolios over one year may outperform them over a five-year time horizon due to changes in the environment, knowledge gains that reconfigure the uncertainty portfolio, or the emergence of opportunity windows.

3.5.2 Organizational and political influences on priority-setting

Recent S&T policy scholarship argues that the decision analytic perspective, which positions research organizations as (more or less) rational assessors of project costs and benefits, only speaks to one part of the story of how R&D priority-setting is conducted at mission-oriented research organizations. This section summarizes the literature addressing the role of organizational and political factors in influencing research governance outcomes.

3.5.2.1 Organizational factors

Several recent studies have highlighted how organizational dynamics play an important role in determining which R&D projects public organizations choose to fund (Brattström and Hellström 2019; Ciarli and Ràfols 2019; Hellström, Jacob, and Sjöö 2017; Cruz-Castro and Sanz-Menéndez 2018; D'Este et al. 2018; Wallace and Ràfols 2018, 2015). Since public organizations are typically slow to respond to changes in their external environment, existing priority-setting practices and routines can become deeply entrenched (Ciarli and Ràfols 2019), and thus may produce R&D portfolios that do not reflect a sober assessment of costs and benefits.

More fundamentally, the assessment of costs and benefits cannot be completely separated from the organizational context from which these assessments emerge. R&D portfolios may be conditioned by a number of organizational factors including: the extent to which the organization is dependent on external sources for funding and expertise (Cruz-Castro and Sanz-Menéndez 2018; Cruz-Castro, Laura, Jonkers, and Sanz-Menéndez 2015), the structure of authority linking researchers and managers (*Ibid.*), the level within the organization where priorities are set (NRC 1995), the incentive structure facing researchers and managers (Stewart 1995) and the extent to which the organizational culture is supportive of mission-oriented research activities (D'Este et al. 2018). In addition to the structural and cultural factors associated with the organizational environment, the S&T policy literature also highlights a number of epistemic or cognitive factors that also skew organizations towards certain types of R&D projects, including the personal values of the researchers (Stewart 1995) and how the conceptual distinction between basic and applied research has considerable influence on R&D decisions (Wallace and Ràfols 2015).

3.5.2.2 Political factors

While organizational factors describe the internal dynamics *within* organizations that influence R&D priority-setting outcomes—or what can be called endogenous factors—political factors can be understood as exogenous influences on priority-setting outcomes that come from *outside* of the organization. Compared to organizational factors, political factors have received far less attention in the literature. Brattström and Hellström (2019, pp. 243) discuss how organizations may tweak their R&D goals "in the direction of political and current social challenges," while Hellström et al. (2017) describe how bargaining over R&D priorities sometimes occurs between levels of government.

From a decision analytic perspective, it seems reasonable to be suspicious of exogenous political factors as they could push organizations away from analytically rigorous CBA and towards the narrow or short-sighted interests of politicians (Martino 1992). However, Stewart (1995, pp. 121) suggests that while "politics is often thought to imply a capricious or arbitrary cast to the direction of research," the contensious bargaining process that often goes on between policy makers and research managers may be necessary to approximate the community's "true" preferences. Neither researchers nor the organizations they work for are perfect arbiters of social value (Neff 2014)—and thus, the input and influence of political actors may ultimately benefit mission-oriented organizations by clarifying their mission.

3.5.3 **Analytical framework**

An effective framework for comprehensively describing and explaining organizational R&D priority-setting must combine organizational and political factors with the typical decision analytic tasks performed by managers at public research organizations. However, it is important to emphasize that organizational and political factors do not simply come into play at the very end of the priority-setting process, once all the decision analysis tasks have been completed. Instead, the influence of organizational and political factors can be seen at every stage of the prioritization process from problem definition to project selection.

The following section describes how both endogenous and exogenous pressures on the NASA planetary defense mission influence the framing of the organization's goals, estimates of the relative benefits produced by candidate R&D projects, as well as implementation decisions like the timing and staffing of projects. The integration of organizational and political factors within every stage of the priority-setting process is reflected in the analytical framework (Fig. 9).

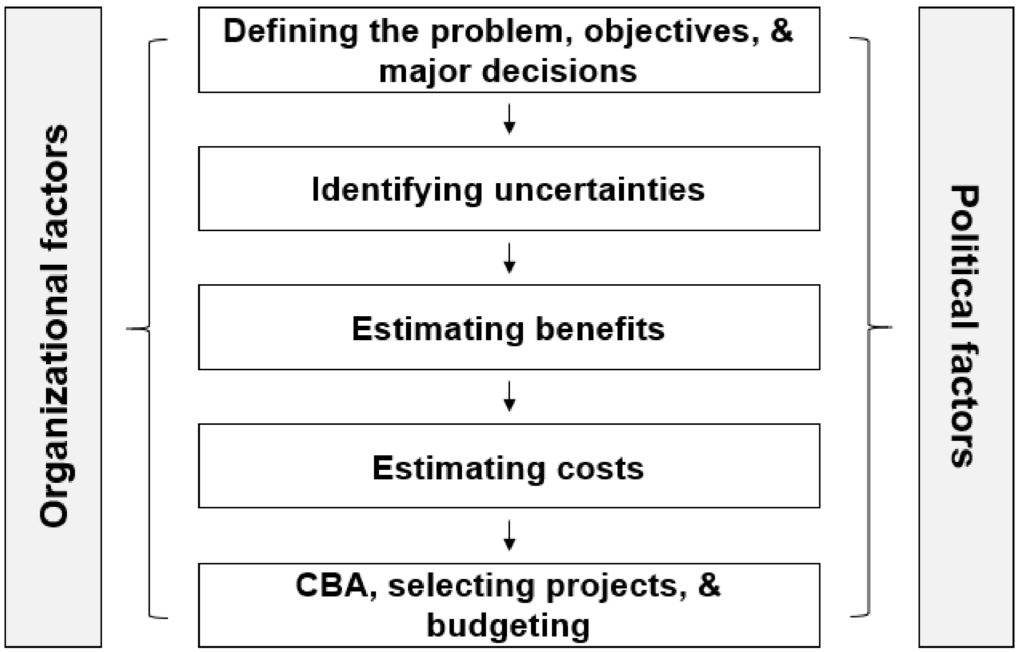

Figure 9: Analytical framework

3.6 Case study: planetary defense R&D priority-setting at NASA

3.6.1 Methodology

A multi-method research design was used to answer two core research questions:

1. What are NASA's current planetary defense R&D priorities and how do they compare to the priorities of the planetary defense expert community?
2. What factors explain planetary defense R&D priority-setting outcomes at NASA?

NASA's current R&D priorities can be easily gleaned from its most recent budget requests and allocations. Two strategies were used to identify the priorities of the broader planetary defense expert community: a bibliometric analysis of the planetary defense literature and a survey. First, I conducted a systematic review and bibliometric analysis of the planetary defense literature using the SCOPUS database. The search was conducted with the assumption that the publication record constitutes an approximation of "revealed priorities" (Ciarli and Ràfols 2019). Relevant peer-reviewed scholarship and conference proceedings published since 2016[16] containing the keywords "planetary defense," "NEO impact," "asteroid impact," or "comet impact" were compiled into a database of 237 sources. A preliminary analysis was conducted to eliminate irrelevant material that did not address the risk of future NEO impacts with the Earth, such as scholarship from the geological and geobiological literature. A total of 197 sources met the bounding criteria.

A notable problem with this database is the glaring underrepresentation of NEO detection and characterization research. The detection and characterization of NEOs is widely considered the most important and active research area for planetary defense.[17] The lack of detection and characterization research picked up by the search terms can be attributed mainly to disciplinary conventions in the fields of astronomy and astrophysics. Most of this scholarship addressing the orbits and physical properties of NEOs does not directly reference the potential of Earth impacts, nor does it widely employ the term planetary defense. Further, NEO discoveries are typically posted in databases like those managed by the Minor Planet Center and often do not appear in academic journals. So, while the bibliometric analysis usefully identified a range of research topics in planetary defense, a different strategy for identifying the current R&D priorities in the planetary defense community was necessary.

To assess the priorities of the broader planetary defense expert community, I also conducted a survey at the 2019 Planetary Defense Conference organized by the International Academy of Astronautics and sponsored by NASA, which is the largest (bi-annual) conference focusing specifically on planetary defense. The survey (N=30) asked conference attendees from government, academia, and the private sector, whose professional activities significantly engage with topics related to planetary defense,[18] to rank the main categories of planetary defense R&D according to their importance for reducing the risk

[16] Literature published before 2016 was excluded from the analysis because the goal was to identify current R&D priorities and perceptions of the value of R&D projects are likely to change over time. The period of 2016-present was selected because it coincides with the existence of the NASA PDCO.
[17] This claim is supported by survey results and interviews with senior NASA research managers.
[18] Survey responses were only considered if the participant indicated that 25% or more of their professional activities engaged with planetary defense.

associated with NEO impacts, as well as their relative cost. In addition to ranking the high-level categories, the survey also asked participants to assess the expected benefits and costs produced by more specific R&D projects in each category.

Semi-structured interviews (N=7) were conducted with all five senior research managers[19] at the NASA Planetary Defense Coordination Office (PDCO), as well as a senior research manager at the NASA-funded Center for NEO Studies (CNEOS) and a member of the Small Bodies Assessment Group (SBAG) Steering Committee. A condensed version of the survey was also administered to interviewees to determine the extent to which their perceived R&D priorities reflected those of the broader community.

3.6.2 **Background**

3.6.2.1 What is planetary defense?

Planetary defense describes activities addressing the risk of large NEOs impacting the Earth. Planetary defense R&D can be grouped into the following five broad categories:

1. detection and characterization of NEOs
2. assessment of impact probabilities
3. assessment of impact and post-impact consequences
4. development of mitigation and adaptation strategies and technologies
5. coordination, rule-making, and public communication

Detection and characterization activities are performed by astronomers using powerful telescopes positioned in observatories around the world, as well as instruments located in space. The assessment of impact probabilities and the consequences associated with various impacts are largely conducted using statistical and mathematical modeling techniques. Mitigation strategies describe possible actions that could be taken to reduce the likelihood of an impact or decrease the magnitude of its consequences (i.e. reduce the number of casualties). Potential deflection techniques include kinetic impactors (where an unmanned spacecraft is launched into an NEO, altering its velocity so it misses the Earth), gravity tractors (where a satellite orbits around an NEO, gradually altering its orbit), and nuclear detonations (where a nuclear device is detonated near the surface of an NEO causing the fragmentation of the object and/or a large, instantaneous change in its trajectory) (NRC 2010). Lastly, R&D on effective coordination, rule-

[19] Senior research manager denotes employees at the position of Program Executive or above.

making, and public communication activities focuses on reducing uncertainty around how to coordinate an effective global response and ensure rapid, legitimate, and high-quality decision-making.

3.6.2.2 Why planetary defense?

Planetary defense is an ideal case for testing the analytical framework and conducting a preliminary analysis of R&D priority-setting for mission-oriented organizations addressing GCRs because the R&D funding ecosystem for planetary defense is significantly less complex than the far more decentralized R&D funding ecosystems of other GCRs. The R&D funding ecosystem for planetary defense is dominated by just two focal organizations: NASA and the European Space Agency (ESA). While many small international, non-governmental, and academic institutions also make important R&D and R&D funding contributions like the International Asteroid Warning Network, the Space Missions Planning Advisory Group, and the United Nations Office for Outer Space Affairs (Pelton 2015), a significant portion of all global planetary defense R&D is funded, directed, or executed by NASA and ESA (Dreier 2019).[20]

NASA and its network of partner observatories are responsible for 98% of all NEO discoveries to date (Landis and Johnson 2019), while ESA has also emerged as a major funder of planetary defense R&D. Since 2009, ESA has invested between 3 to 9 million euros annually to planetary defense activities and recently awarded a 129 million euro contract to the German space company OHB to develop Hera—ESA's contribution to the joint NASA-ESA double asteroid redirection test scheduled for 2021/2022 (Clark 2020). The concentration of major planetary defense R&D funding decisions within these two organizations means that an in-depth analysis of just one of these organization's R&D priority-setting activities can capture a significant part of the overall planetary defense R&D landscape. Furthermore, both NASA and ESA's R&D priority-setting activities cover the entire planetary defense uncertainty portfolio.

This section applies the analytical framework to an in-depth case study of planetary defense R&D priority-setting at NASA specifically. While the case study highlights some of the interactions between

[20] Barrett (2007) argues that the relatively sparse planetary defense governance landscape stems from the nature of the NEO impact risk itself. Barrett describes planetary defense as a "single best effort" problem where it only takes one actor to supply a global public good that benefits everyone, disincentivizing participation from other actors. However, recent planetary defense simulations and tabletop exercises have demonstrated that planetary defense decisions made by one country could pose significant risks for others, especially if they involve the deployment of nuclear weapons (CNEOS 2019). These experiences highlight the importance of broader international participation in planetary defense.

NASA and ESA in the development of the double asteroid redirection test, the focus of the study is on the internal R&D priority-setting activities of NASA's planetary defense mission and the implications for other mission-oriented organizations. However, scholarship that maps out the broader planetary defense R&D ecosystem and analyzes the impact of interorganizational dynamics on R&D priority-setting are needed to provide a more comprehensive picture of the evolving planetary defense governance regime.

3.6.3 Planetary defense R&D priorities at NASA

NASA's NEO Observations program began in 1998 after the U.S. Congress directed NASA to discover 90% of all NEOs with a diameter of 1 km or larger—a goal that was expanded in 2005 to NEOs with a diameter of 140 m or larger.[21] Responding to a recommendation from a 2014 report by the NASA Office of Inspector General, NASA established the Planetary Defense Coordination Office (PDCO) in 2016 to manage all of NASA's detection, assessment, mitigation, and interagency coordination activities, including the management of the planetary defense R&D portfolio and annual budget.

NASA's planetary defense budget increased from a mere 3.3 million USD in 2008 to 157 million USD in 2019, while its share of the overall NASA budget grew from 0.02% to 0.73% over that period (Dreier 2019). As reflected by the 2019 budget, NASA's primary R&D priority for planetary defense is the Double Asteroid Redirection Test (DART) mission, which accounts for approximately 65%[22] of the 157 million USD appropriated for planetary defense, which is expected to decrease incrementally each year until the mission is completed in 2022. The DART mission, which is directed by NASA and carried out by the Applied Physics Laboratory (APL) at Johns Hopkins University, is a test of the kinetic impactor deflection method. The DART spacecraft is scheduled to launch in late-2021 and will be intentionally crashed into a small satellite asteroid orbiting around a larger asteroid called Didymos as it makes a close approach with Earth in 2022. The mission is funded entirely by an additional influx of funding on top of the "base" planetary defense budget.

The second largest line in the 2019 budget is for detection and characterization activities coordinated through the NEO Observations Program, which draws about 25% of the total budget (or 90% of the non-DART planetary defense budget). The remaining budget is split somewhat evenly between R&D programs and projects addressing impact and post-impact consequences, impact probability assessment,

[21] More detailed discussions of the history of planetary defense at NASA can be found in Baum (2019) and Landis & Johnson (2019).
[22] Estimates of the budget requested and allocated to the DART mission range from 62% (Dreier 2019; NASA 2018a) to 67% (Source: interviews).

and interagency and international coordination activities. High-level research priorities reflected by the 2019 budget are shown in Table 4.

Table 4: NASA 2019 planetary defense R&D priorities (source: interviews)

High-level research area	Proportion of FY2019 budget
Mitigation & adaptation strategies and technologies	62-67%
Detection & characterization	25%
Impact and post-impact consequences	5%
Impact probability assessment	~1%
Coordination, rule-making, & public communication	~1%
Miscellaneous	~1%

However, survey results show a slightly different picture of the perceived R&D priorities of senior research managers within the PDCO than those reflected in the budget. All interviewees indicated that detection and characterization is (by far) the most important R&D area for decreasing the risk of an NEO impact. In other words, they were unanimous in their belief that going forward, NEO observations, follow-ups and characterizations will provide the greatest benefit or social value. Mitigation activities and impact probability assessment both received votes as the second most important R&D area going forward, while impact and post-impact consequence analysis is seen as the least important.

For NASA employees, the perceived importance of detection and characterization R&D is linked to the proposed 500 million USD[23] space-based infrared telescope project called NEOCam, which is specifically designed to address the goal of discovering 90% of the NEO population of 140 m and larger. NEOCam is currently categorized as an extended Phase A study, which means that it is in the "pre-formulation or formulation stages" of its life cycle (NASA 2019a) but prior to 2019, it was seen internally as the main competitor for funding with DART. The preference within the PDCO for NEOCam over DART is apparent, with most interviewees sharing the sentiment expressed by one research manager that "you've got to find them before you figure out what to do about them."

The opinions of NASA employees closely reflect those of the broader planetary defense expert community. 87% of survey respondents indicated that detection and characterization was the most important R&D area going forward, while 13% indicated it was the second-most important. By

[23] Source: (Fernholz 2019)

comparison, mitigation and adaptation (which would include a kinetic impactor test) was ranked, on average, the third-most important R&D area but the most expensive. 83% of participants saw “modeling, developing, and testing kinetic impactor strategies and technologies” as either “very expensive” or “extremely expensive” relative to other R&D activities. Survey results also indicate that asteroid impact probability assessment is believed to have one of the highest “returns on investment.” Participants, on average, ranked asteroid impact probability assessment as being the second-most important and the second cheapest of the 11 R&D areas. Survey results are summarized in Figure 10.

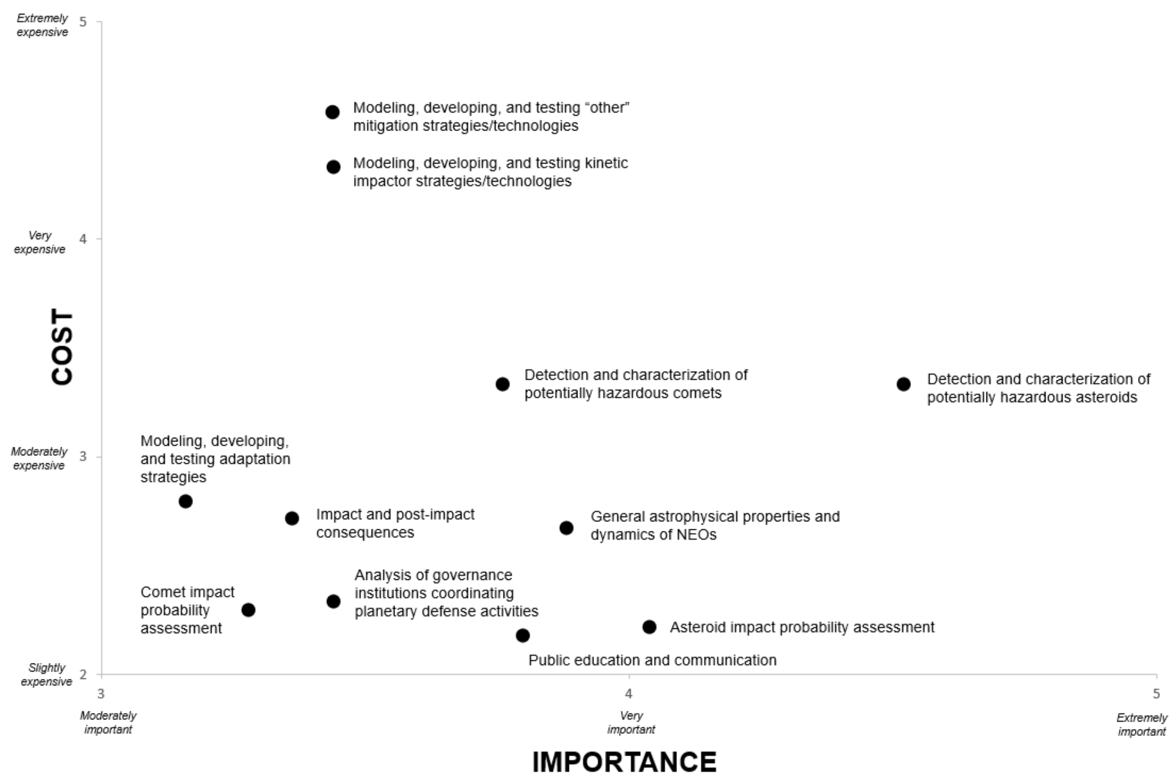

Figure 10: Average assessment of importance and cost of planetary defense R&D activities. (source: survey)

3.6.4 The R&D priority-setting process at the PDCO

3.6.4.1 Overview

The NASA planetary defense budget is administered by the PDCO, which is located within the Planetary Science Division of NASA’s Science Mission Directorate (SMD). The PDCO allocates its base funding through two channels. First, about half of the PDCO’s base funding supports programs and projects at arms-length organizations like CNEOS and other contractors. For example, the NASA Infrared Telescope Facility (IRTF) is largely funded by NASA but operated and managed by the University of Hawaii. Funding for projects at these arms-length organizations is not openly competed through a public grant process since these organizations offer unique and mission-critical capabilities—but they are still subject to peer review and regular evaluation. As one interviewee noted: "Until we have something better, we don't want to lose those capabilities. That would violate the first rule of wing walking: never let go of something until you've got a hold of something else."

The other half of base funding is administered through NASA's public grant competition called Research Opportunities in Space and Earth Science (ROSES), which is the primary focus of this analysis. ROSES is an annual competition that awards one to three-year grants to research teams or individual researchers located at universities, observatories, and private firms. Project proposals are evaluated through peer review according to their perceived merit, their cost and feasibility, and their relevance to the criteria and priorities expressed in the planetary defense appendix of the ROSES call prepared by senior research managers at the PDCO.

This analysis focuses on the ROSES competition process because, unlike the less structured process through which the PDCO renews ongoing funding for key observatories, the ROSES competition employs standardized (and thus more easily described) processes for measuring and comparing the costs and benefits of new projects. While outside of the scope of this study, an analysis of how legacy observatories are evaluated at NASA would help provide a more complete picture of the R&D ecosystem around planetary defense at NASA. An organizational diagram of this ecosystem is provided in Figure11.

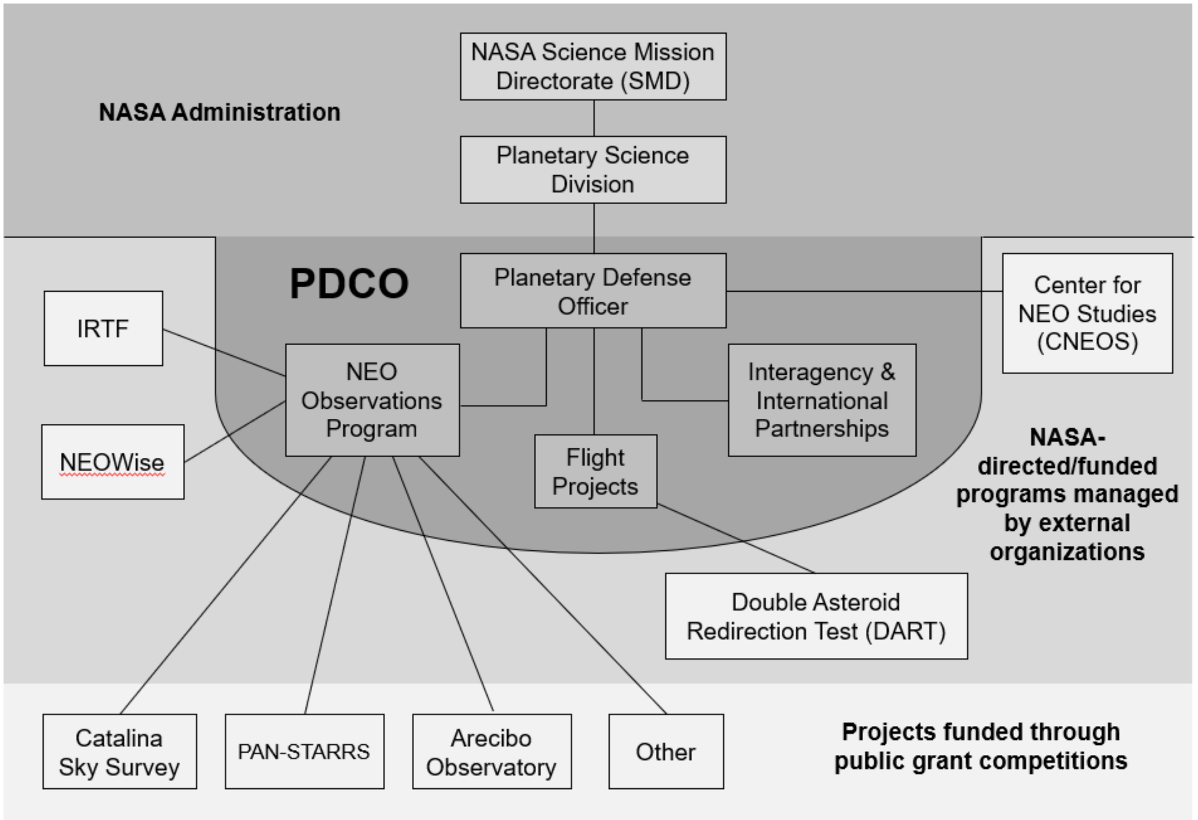

Figure 11: PDCO organizational chart and funding system (source: adapted from NASA 2019b)

Key decision makers in the PDCO for R&D priority-setting include the Planetary Defense Officer (i.e. the PDCO director), as well as program executives, program managers, and program scientists located at NASA headquarters. Senior research managers at NASA-affiliated external organizations also exert influence on R&D strategy and budgeting decisions through regular interactions with PDCO staff and the submission of budget requests for their individual programs. Ultimately, final authority for approving PDCO's budget proposals and making ROSES grant allocation decisions, rests with the Planetary Science Division Director within the SMD.

The PDCO also solicits opinions from advisory bodies like the Small Bodies Assessment Group (SBAG), which is made up of experts on NEOs from academia and industry and presents "findings" that are widely read by PDCO staff. Research managers at the PDCO also lean heavily on what one interviewee described as "what-NASA-should-do-next reports" like the National Research Council's (NRC) decadal survey (2010) and two Science Definition Team (SDT) reports (NEO SDT, 2003, 2017). The SDT reports were chartered by NASA to study the feasibility of extending the search for NEOs to objects smaller than 1 km in diameter. The reports present a series of detailed cost-benefit analyses of candidate NEO detection system configurations, including several combinations of space and ground-based observatories. Notably, the 2017 SDT report offers strong support for investing in a space-based telescope like NEOCam. All interviewees indicated that the SDT reports have been highly influential in shaping planetary defense R&D priorities, with one interviewee stating: “had [the SDT] said a space-based observatory would be merely a ‘nice to have,’ we wouldn't have nearly the grounds with which to push forward with it that we do now.”

In the following analysis, I consider these formal advisory bodies, as well as arms-length organizations like CNEOS, to be organizational (i.e. endogenous) factors influencing R&D governance, while I consider Congress, the White House, and the broader academic research community to be political (i.e. exogenous) factors. This distinction is consistent with how interviewees’ described the boundary line separating NASA’s formal advisory process from influences they consider to be external to their internal R&D process.

3.6.4.2 Defining the problem, objectives, and major decisions

As I argue in Section 3.5.1.1, problem formulation or risk identification is an intersubjective process where there is broad agreement on a causal story describing how a system that produces social value (like human welfare or continued existence) may be compromised. There is a large intersubjective consensus within the PDCO and the broader planetary defense community on the nature of the NEO impact problem. When asked about the fundamental problem addressed by planetary defense governance, interviewees gave remarkably similar answers, pointing to the robust geological record as evidence that the question is not *if* an NEO posing large or catastrophic consequences for humanity will impact the Earth but *when*—and whether we will be capable of stopping it.

Key objectives for planetary defense are described by Lee et al. (2014), including minimizing human mortality and injury, as well as damage to critical infrastructure, property, and ecosystems. Framed in operational terms, one interviewee suggested that planetary defense boils down to three central objectives: “Our first goal is to find them and characterize them. Our second goal is to know what to do about them. And our third goal is to build the policies and plans in the government so that we can respond effectively.”

Several major decision points emerge from the problem definition and key objectives, such as determining if an NEO poses an impact hazard or not, if it should be deflected, when to deflect it, what type of deflection technology and strategy to use, what disaster management strategy to use, and which actors are responsible for executing these actions.

It is important to note that the PDCO research managers controlling the R&D priority-setting process do not have a monopoly on defining the problem and identifying objectives and key decision points. Many of these ideas, including the idea of developing and testing the deflection mitigation technique, did not emerge within NASA but from the academic community studying NEOs and planetary defense. Therefore, exogenous (or political) factors have played a key role in shaping how PDCO research managers understand the problem, objectives, and key decision points.

Another source of external influence on this stage in the R&D prioritization process is the SDT reports, which produced the 140 m threshold that NASA uses to define "dangerous" NEOs. PDCO staff admit that this threshold is somewhat arbitrary. For instance, an asteroid estimated to have a diameter of just 18 m exploded over the city of Chelyabinsk, Russia in 2005 (NASA 2013), and while the event did not lead to any casualties, one NASA research manager pointed to a number of fortunate factors, such as the fact that the event took place immediately after the morning commute when most individuals were already indoors, that prevented the incident from being significantly more harmful. Despite the uncertainty around whether the search should be expanded to smaller objects, most PDCO staff were content with the SDT report's 140 m threshold.

3.6.4.3 Identifying uncertainties and candidate projects

The identification of key uncertainties is not conducted through a systematic process at the PDCO. However, the close interaction between the PDCO and the external planetary defense research community means that the planetary defense uncertainty portfolio is informally tracked and managed by academic journals like *Acta Astronautica* and *Space Policy*, while also being regularly summarized and updated at planetary defense conferences. The main impression from interviews with senior research managers is that high-level uncertainties are widely known by all PDCO staff, while more detailed topic-specific uncertainties are "owned" by program executives working on those specific issues.

Interviewees generally agreed with the five broad categories of planetary defense R&D they were asked to rank, but also highlighted that there are many overlaps and interdependencies between them—particularly between observation, characterization, and impact assessment research. For instance, impact assessment research has almost no value without new observations and accurate characterizations. Bearing in mind that the boundaries between these categories are somewhat permeable, several examples of key uncertainties from each R&D category are highlighted in Table 5 based on interview responses and a review of the planetary defense literature.

Table 5: Planetary defense uncertainty portfolio

R&D category	Examples of key uncertainties
Detection and characterization of NEOs	• What is the total population of NEOs >140 m diameter? • What is the population of "lost asteroids"? • What is the orbit, diameter, mass, spin, and composition of different NEOs? • Are current observation capabilities sufficient for effective characterization and follow-up?
Impact probability assessment	• Does an NEO >140 m diameter pose an impact hazard this century? • What is the historical impact rate of NEOs >140 m diameter? • How accurate is the historical impact rate? • To what extent is the historical impact rate an accurate guide of the near-future impact rate?
Impact and post-impact consequence assessment	• To what extent do uncertain parameters like: impact velocity, diameter, mass, density, strength, and entry angle… lead to uncertain impacts like: airburst altitude, thermal radiation and blast damage, tsunami creation, and global climatic effects? • What are the likely impacts at different geographical locations? • What is the relationship between impacts and human mortality, injuries, damage to critical infrastructure, damage to ecosystems, and damage to property? • To what extent do value of statistical life (VSL) calculations accurately capture the costs associated with human mortality?
Mitigation and adaptation	• Which deflection technology/strategy is the most (cost-)effective? • Have deflection technologies/strategies been sufficiently demonstrated and evaluated? • In the event of mission failure, is there sufficient time for a second effort? • Is a nuclear detonation strategy politically, technologically, or financially feasible? • When do adaptation strategies need to be initiated?
Coordination, rule-making, and public communication	• When should the public be informed of an impending NEO impact? • To what extent are relevant actors and organizations sufficiently coordinated and prepared for a deflection mission or disaster response? • What policies and procedures are necessary to effectively execute a deflection mission or respond to an impact in real-time? • To what extent can national governments act unilaterally to mitigate the risk of an NEO impact?

The process for identifying candidate projects centers around the annual ROSES call, which outlines the key gaps in NASA's R&D portfolio. PDCO staff note that the call changes very little year-to-year, with the main focus on R&D projects that "promise a sustained, productive search for NEOs and/or obtain follow-up observations of sufficient astrometric precision to allow the accurate prediction of the trajectories of all discovered objects" (NASA, 2018b, pp. C.6-2). Occasionally, emphasis is added to a particular research area. In 2018, the call included text about characterization studies that specifically support mitigation actions (*Ibid*). Candidate projects can also emerge from NASA itself or its arms-length partner organizations. For example, a research manager at CNEOS described how they communicate new project ideas directly to the Planetary Defense Officer at the PDCO. NEOCam was proposed by JPL through the Discovery Program—a special competition for space missions held by NASA every few years.

The PDCO has stressed keeping their search for candidate R&D projects as broad as possible. One research manager noted that the call is intentionally kept somewhat vague in order to "leave some room open for creative ideas," while another suggested that the PDCO's strategic priorities are intentionally omitted or only hinted at in the ROSES call "so it doesn't seem like NASA is shutting down new innovation." However, one political factor that clearly constrains which candidate projects get considered is the policy prohibiting non-U.S. organizations, collaborators or subcontracts from receiving NASA funding (NASA 2018b).

3.6.4.4 Estimating benefits

The assessment of the benefits or social value produced by candidate R&D projects is largely conducted through peer review. Each ROSES proposal is adjudicated by panels made up of three reviewers from NASA and two to four external reviewers (often from academia). Reviewers are tasked with assessing the merit and relevance of each project and scoring them on a scale from one (poor) to five (excellent) and justifying their scores with a written narrative. One interviewee noted that reviewers are warned to "watch their adverbs and adjectives" in the narrative to protect the review process from criticism for being overly subjective.

However, peer review is inescapably (inter)subjective. ROSES reviewers are first tasked with assessing the relevance of a proposed project to the specific language in the call—which PDCO research managers admit is kept intentionally imprecise. Next, reviewers must estimate the merit of the project, which can be translated in decision analytic terms as how significantly a particular project decreases important uncertainties impeding planetary defense decision-making. However, when asked how they interpret and measure merit, none of the interviewees made the connection between the merit of individual R&D projects and uncertainty reductions and improvements in decision-making. Instead, they pointed to proxy measurements of social value such as the "scientific value" or "importance" of proposed projects.

Another benefit estimation process that runs quietly alongside peer review at the PDCO is the quantified benefit estimations in the SDT reports, which PDCO research managers lean on to justify maintaining current observation capabilities and to advocate for a space-based infrared telescope. The 2017 SDT report reflects an uncertainty reduction interpretation of benefit estimation, defining benefit as: "reducing the uncertainties of hazards to life, injuries, and property/infrastructure damage resulting from impacts on Earth over a 100-year time horizon" (NEO SDT, 2017).

The SDT reports measure benefits using value of statistical life (VSL) estimates that describe the amount of financial harm produced by the unretired impact hazard (which is determined by historical impact statistics). The 2017 SDT estimates that the residual risk from NEO impacts is approximately 757 million USD per year. Therefore, an observation and characterization capability that would discover all remaining NEOs larger than 140 m and eliminate this risk would produce annual benefits of approximately 757 million USD. While none of the configurations of observatories assessed in the report are capable of completely eliminating the impact hazard, the report estimates that over half of the systems analyzed would produce over 1 billion USD of total benefits accrued over the next 20 years (compared to just 250 million USD if no new capabilities are added) (NEO SDT, 2017).

Interviewees acknowledged the many challenges associated with the *ex ante* estimation of benefits produced by candidate R&D projects. While all PDCO research managers assigned greater value to NEOCam than to DART in a vacuum, some were quick to point out that the intervening factor of timing should perhaps influence the assessment of the benefits produced by the DART mission, with one individual stating:

> All things being equal, a space-based survey capability would have been my choice. However, you also have to take into account the opportunity that nature has given us with the Didymos close approach in 2022. In this business, it's not always about the money available, it's also your opportunities.

Another complicating factor is the interdependencies between projects. For instance, if follow-up capabilities such as the precision tracking of NEOs lags behind detection capabilities, estimates of the NEO impact hazard may become inflated (Reich 2010). One research manager also commented on how the value of the DART mission had been decreased significantly by the ESA's withdrawal from the DART mission in 2016 (ESA later rejoined the mission in 2020), which would perform monitoring and characterization tasks before and after the DART impact. Instead, NASA would be forced to rely on more limited ground-based observations. Highlighting the complementarity between these two projects, the interviewee stated: "Without knowledge of the mass of the moon, we can't have any idea of what the beta factor is.[24] And that is basically the key result of a mitigation demonstration like the DART mission. I think that's very unfortunate."

Perhaps the most difficult task associated with benefit estimation for planetary defense R&D is comparing the value of projects that address different research areas, uncertainties, and key decisions. In order to better understand how PDCO research managers make these sorts of comparisons, interviewees were asked to describe the precise circumstances where they would be willing to prioritize mitigation R&D over observations and characterizations. Most interviewees struggled to describe the precise circumstances—however, one interviewee appealed to specific NEO discovery statistics, stating: "For the >140 m ones, I think I would want to be pushing 90% before I'd be willing to flip and make mitigation the higher priority."

The influence of organizational and political factors on benefit estimations is also significant. First, there is a strong organizational culture both within NASA and US government agencies in general around maintaining perceptions of objectivity and fairness (Martino 1992). This sensitivity underpins the PDCO's stated preference for putting as many candidate projects as possible through the peer review process, which lends legitimacy to priority-setting and budgeting decisions.

Perhaps the most significant form of organizational influence over benefit estimation is the NASA SMD's narrow definition of science (Chapman 1999; Baum 2019). One of the points frequently made by interviewees was that many of the planetary defense R&D projects proposed by the PDCO are not considered "real science" by key decision makers within the SMD and are thus at a considerable disadvantage when competing against projects that *are* considered real science. Interviewees described how the SMD defines science narrowly as "hypothesis-driven inquiry." When asked why planetary defense R&D does not qualify as hypothesis-driven inquiry, one research manager responded, "Well, I guess we *could* say: we think there are 7.2 million asteroids out there, let's go find out. But the response

[24] The beta is the momentum enhancement factor which determines the amount of velocity transferred to the asteroid when it is impacted (Heberling et al. 2017).

is: It's counting. It's not science." In other words, there exists a bias in the NASA administration that inflates the estimated benefits of projects deemed to be science and deflates the estimated benefits of planetary defense projects that do not fit that mold.

Other interviewees were less concerned about expanding the SMD's narrow definition of science but instead argued that planetary defense R&D should not have to compete with science projects at all: "NEOCam is competing with science missions for science dollars, which is a big disadvantage. Planetary defense missions should compete with one another rather than competing with science missions." Expressing a similar sentiment, another research manager commented:

> If it were taken out of science and made independent, even with the modest budget it has, [PDCO] would be better able to make its own decisions, rather than have that sort of interference that's going on. I'm not implying that it's the people, it's just that SMD is oversubscribed. Planetary defense is not quite a square peg—but it does not fit all that well.

3.6.4.5 Estimating costs

The task of estimating the costs of candidate R&D projects is simplified significantly by the peer review process. Organizations submitting ROSES proposals are required to estimate their own costs, including how much time and resources the project requires to deliver on its stated outcomes, as well as potential project risks. The task of ROSES reviewers is simply to assess how realistic those estimates are. One interviewee commented: "The question is: are they going to be able to accomplish what we want them to do? Quite frankly, in most cases they are not asking for enough."

Like with benefit estimation, the SDT reports are highly influential on the perceptions of research managers on the relative costs of different NEO detection systems. The SDT reports use models like parametric cost-estimating relationships and the NASA Instrument Cost Model to estimate the cost of different ground- and space-based detection systems. However, in other research areas like mitigation where there is a lack of analogous systems to use as the basis for cost estimation, these estimates can be unreliable. According to the 2010 NRC decadal survey, "At best, these estimates provide only crude approximations of final costs of pursuing any of these options" (NRC 2010, pp. 97).

Sometimes organizational and political factors interact to complicate the task of accurately estimating the costs of candidate projects. A cost assessment for the DART project was initially conducted by the APL at Johns Hopkins when the project emerged in 2012. However, initial cost estimates were based on how the project was described in the proposal, which positioned it as a Category 3 mission—a classification given to low-cost, high-risk projects that tend to be developed and executed fairly rapidly. However, as DART moved through the pre-formulation and formulation stages of development and received funds from Congress, the project began to receive considerably more attention from the rest of NASA, the Office of Management and Budget (OMB), and Congress. As a result, DART was reclassified as a Category 2 mission, which significantly increased the amount of oversight and cost. While, the project is still set to meet its deadline for a launch in late-2021, one research manager concluded that, "to some extent, DART has been the victim of its own success." Unexpected attention and enthusiasm from outside of the PDCO has led to budget increases that were difficult to anticipate at the outset of the project.

3.6.4.6 Integrating costs and benefits, selecting projects, and budgeting

Combining cost and benefit estimations into an integrated assessment of candidate projects is, to some extent, conducted mathematically through the peer review scoring system. However, reviewers only assess the accuracy of cost estimates, not the costs themselves. PDCO research managers and administrators within the Planetary Science Division do not typically use formal quantitative methods to analyze project benefits and their associated costs. One exception is the cost-benefit ratios of NEO detection systems presented in the SDT reports, which calculate the amount of benefit (measured in USD) produced by each dollar spent. However, for other R&D areas, the CBA tends to be more implicit and is folded into the PDCO's somewhat opaque project selection process.

Once the value of candidate projects is assessed and decision makers have a sense of how much investment is required to produce that value, difficult decisions must be made about how to allocate a finite budget across an overabundance of worthy projects. In addition to estimating the value produced by each project per dollar spent, research managers must also factor in the existing portfolio of projects already being funded. There appears to be wide agreement within the PDCO that its R&D portfolio should touch the entire planetary defense uncertainty portfolio—a sentiment that was reflected by the difficulty that research managers had ranking the importance of planetary defense research areas.

The task of assessing the "fit" of candidate projects within the existing R&D portfolio is conducted, to some extent, by the PDCO research managers who analyze the peer review results and prepare a series of recommendations for the Planetary Defense Officer. However, the recommendations are eventually passed up to the Research Director of the Planetary Science Division who has ultimate authority over all planetary defense funding decisions. Described by one research manager, once recommendations are passed up to the Research Director and eventually the White House and the OMB, PDCO research managers continue to try to influence project selection:

> You've got to sell your boss. You've got to sell the OMB. And then you've got to sell Congress—and that's an iterative process too. And all of that can be affected by outside forces. None of it is as tidy as we ever want it to be.

As this comment hints, exogenous political influences on R&D decision-making at the PDCO are most acute at the project selection stage where powerful actors can swoop in and impose their own priorities on planetary defense budget allocation. This dynamic played out in the 2019 budget with the decision to fund the DART mission. While the 97 million USD appropriated by Congress for the DART mission in 2019 did not directly displace funding earmarked for other R&D projects, the general opinion amongst PDCO research managers is that the decision to fund the DART mission effectively blocked any hope of funding NEOCam in the next few years. The insertion of the DART mission into the planetary defense budget is generally attributed to Congress. According to one research manager, the preference for DART over NEOCam is largely the result of the political influence of the APL at Johns Hopkins operating through the Senator from Maryland. Another interviewee summarized the influence of political actors on NASA generally:

> Experience has taught me that Congress is hugely influential. I think that while most NASA decisions are probably *not* influenced by Congress, if Congress wants to influence a decision, they can. The White House can also have a huge influence on NASA. So, I wouldn't say that NASA is at their whim—but it's very much subject to pressures from the White House and Congress because ultimately, it's Congress that pays the bills and NASA is a part of the executive branch of government.

While the NASA budget is certainly not immune to pork barrel politics, most PDCO research managers are quick to point out the double-edged sword of political influence on the R&D priority-setting process. For instance, there was certainly no guarantee that if Congress did not appropriate funds for DART it would have funded NEOCam instead. This point was not lost on one research manager who reflected: "I'm caught in the uncomfortable position of complaining about this big budget increase that we have been given because it doesn't accomplish everything we thought we needed to do for the program."

Interviewees also reflected on the nature of the PDCO as a mission-oriented research organization where their mission is ultimately to protect the US public and people around the world, with one research manager acknowledging: "Congressional Representatives would not be doing their job unless they were faithfully representing their constituents and making sure they were getting their fair piece of the pie."

3.6.4.7 Conclusions and limitations

This case study provides an in-depth analysis of R&D priority setting at NASA's planetary defense mission and illustrates how organizational and political factors interact with every decision analytic task carried out by research managers at the PDCO. Furthermore, this close look at NASA not only sheds light on the priority-setting processes of one organization but also a significant portion of the overall planetary defense R&D landscape. However, NASA is one of two focal organizations in the planetary defense R&D landscape. A more thorough analysis of the overall planetary defense R&D landscape would necessarily require a similar analysis of the ESA and a greater focus on the interactions between ESA and NASA. The case study's focus on the ROSES competition also ignores the less formal processes for evaluating and renewing contracts with legacy organizations, which account for half of NASA's base funding for planetary defense. One might expect organizational and political factors to play an even greater role in influencing these outcomes—however, more research is needed to compare the formal and informal R&D priority-setting processes used by NASA and mission-oriented organizations more generally.

3.7 Conclusion: lessons for GCR research governance

The final task of this article is to consider the extent to which planetary defense R&D priority-setting at NASA can inform the analysis of other mission-oriented research organizations addressing GCRs. Like planetary defense, investment decisions for climate change R&D or disease outbreak R&D could have enormous implications for our ability to adequately understand and respond to these risks in time. As discussed in Section 3.6.2, planetary defense is an unusually tidy R&D ecosystem with a large amount of activity funded or directed by two focal organizations whose R&D portfolios touch all corners of the planetary defense uncertainty portfolio. However, the NASA case still offers a number of useful lessons for understanding and improving R&D priority-setting for other GCRs.

One of the main lessons that can be applied from the planetary defense case to other GCRs is the necessity of developing formal systems to contend with the unavoidably messy task of R&D benefit estimation. The PDCO uses a combination of qualitative and quantitative approaches to estimate benefits. Peer review scores for the merit and relevance of candidate projects are used as crude approximations of their relative social value, while more rigorous quantitative estimates of the NEO impact hazard calculated in the SDT reports are also used by PDCO staff to justify their decisions.

While using a combination of qualitative and quantitative benefit estimation methods may be useful for other GCRs with robust historical data like supervolcanoes, the vast majority of GCRs are unprecedented or lack adequate historical data. Comparing planetary defense to the risk of nuclear war, Baum (2019) writes: “The statistics of NEO collision have a relatively strong empirical basis, whereas the risk of nuclear war depends on ambiguous factors such as the tendency for national leadership to launch nuclear weapons.” Emerging risks like AI accidents or bioterrorism where we do not even know the precise form a risk event might take are the least amenable to quantitative risk analysis. Therefore, assessments of these types of risks tend to rely on intersubjective expert elicitation methods (e.g. Grace et al. 2018; Sandberg and Bostrom 2008). The PDCO’s experience with peer review is informative for these types of risks. While certainly slower and more expensive than evaluating R&D candidate projects without external reviewers and formalized scoring systems, research managers at the PDCO emphasize how peer review offers a crucial layer of (perceived) objectivity and protects the organization from accusations of bias or conflict of interest.

Another lesson from the planetary defense case is the importance of convening relevant knowledge centres across government, academia and the private sector, thereby fostering a research community that can “funnel” relevant R&D towards focal organizations like NASA. The bibliometric analysis performed in this study encountered the problem that a significant portion of relevant detection and characterization research does not explicitly self-identify as planetary defense R&D. NASA-sponsored planetary defense conferences bring together a large number of astronomers working on NEO observations that do not necessarily identify as “planetary defense researchers”—however, their research findings are critical to estimating the NEO impact hazard. Similarly, many computer scientists are unaware of the risk of catastrophic AI accidents but may possess expertise that is highly relevant to efforts to address them. The institutionalization of regular consultations with advisory groups like SBAG also helps to develop ties with “mission-adjacent” organizations and is a model that can be adopted by organizations addressing other GCRs.

Perhaps the key lesson that should be taken from the case study is the context-dependent nature of organizational priority-setting. Even with few interorganizational interactions, R&D decision-making at NASA is still constrained by organizational and political factors during each stage of the decision analysis cycle. At times these factors can be decisive, such as the insertion of the DART project into the planetary defense budget, displacing alternative projects preferred by research managers. The NASA case shows that even for relatively insulated organizations, careful attention should be paid to the host of organizational and political factors overlooked by much of the priority-setting literature.

Chapter 4

Article #3: Confident, likely, or both? The implementation of the uncertainty language framework in IPCC special reports

4.1 Preface

This article explores the characterization and communication of uncertainty by the Intergovernmental Panel on Climate Change (IPCC), which is tasked with providing authoritative knowledge on the causes and impacts of – and responses to – climate change to the signatories of the United Nations Framework Convention on Climate Change. The IPCC is widely considered the foremost authority on the state of climate change knowledge, which it communicates through the publication of assessment reports that are prepared by hundreds of the world's foremost natural and social scientists. These assessment reports communicate both what we know and what we do not know about climate change. As I argue in Article #1, most knowledge claims fall somewhere between the poles of "true" and "false," and thus the task for IPCC lead authors is to communicate where exactly they fall based on the current state of scientific knowledge.

In order to accomplish this task, IPCC lead authors use a series of calibrated uncertainty terms to qualify uncertain knowledge claims, which are prescribed by the IPCC's uncertainty language framework. Almost all of the criticism and analysis of this framework has appeared in the journal *Climatic Change*, which also published a version of this article in May 2020 (Janzwood 2020).

This article is informed by the discussion of uncertainty "levels" presented in Article #1. In that article, I tether uncertainty levels like "shallow" and "deep" uncertainty to different *uncertainty structures,* which describe situations when it is more and less appropriate to quantify uncertainty. However, according to the subjectivist Bayesian interpretation of uncertainty, probability values can technically be assigned to any uncertainty situation (even if it may be arbitrary or misleading to do so). This perspective can be contrasted with the frequentist notion of probability where probability can only be quantified if there exists a robust and homogenous set of historical trials or model runs. This distinction is highly relevant to the IPCC's uncertainty language framework, where great effort has been made to distinguish between uncertainty situations that can and cannot be described with probability values.

The IPCC's framework prescribes qualitative confidence terms like "medium confidence" for communicating uncertainties that lack a quantitative basis (i.e. frequencies), while likelihood terms like "very likely" that are linked to specific probability intervals (e.g. 90-100%) are used to describe the outputs of quantitative (and largely frequentist) analyses and modeling studies. However, this distinction is the source of tremendous confusion in the literature (and amongst IPCC lead authors themselves). First, it is not clear why confidence levels cannot be expressed probabilistically (which would be the interpretation that aligns with the Bayesian or confidence-deficit interpretation of probability that I present in Article #1). Second, it is not clear whether confidence alone, likelihood alone, or both terms constitute expressions of the "validity" or "truthfulness" of the knowledge claim. The sources of these confusions are explored in the article.

While Article #2 investigates decision support relationships at NASA, where experts and policy makers are located within the same organization (and the line between the two is sometimes blurred), the experts contributing to IPCC assessment reports are clearly detached from their target audience and have a more consultative function. Further, while many top NASA administrators and decision makers have a background in science or engineering, many key decision makers within national governments with respect to climate change policy cannot be said to possess an equivalent level of background knowledge. Therefore, the importance of clear and accurate uncertainty communication at the IPCC is magnified.

4.2 Abstract

The uncertainty language framework used by the Intergovernmental Panel on Climate Change (IPCC) is designed to encourage the consistent characterization and communication of uncertainty between chapters, working groups, and reports. However, the framework has not been updated since 2010, despite criticism that it was applied inconsistently in the Fifth Assessment Report (AR5) and that the distinctions between the framework's three language scales remain unclear. This article presents a mixed methods analysis of the application and underlying interpretation of the uncertainty language framework by IPCC authors in the three special reports published since AR5. First, I present an analysis of uncertainty language term usage in three recent special reports: Global Warming of 1.5°C (SR15), Climate Change and Land (SRCCL), and The Ocean and Cryosphere in a Changing Climate (SROCC). The language usage analysis highlights how many of the trends identified in previous reports – like the significant increase in the use of confidence terms – have carried forward into the special reports. These observed trends, along with ongoing debates in the literature on how to interpret the framework's three language scales, inform an analysis of IPCC author interviews based on their experiences interpreting and

implementing the framework. Lastly, I propose several recommendations for clarifying the IPCC uncertainty language framework to address persistent sources of confusion highlighted by the authors.

4.3 Introduction

The Intergovernmental Panel on Climate Change (IPCC) is tasked with providing comprehensive assessments of the state of knowledge on the causes and impacts of – and responses to – climate change. In order to effectively communicate expert judgements on thousands of policy-relevant knowledge claims made in each report, the IPCC has implemented a system of calibrated uncertainty language that is designed to encourage the consistent characterization and communication of uncertainty between chapters, working groups, and reports. The IPCC uncertainty language framework has been revised or clarified before each of the last three major assessment reports. However, the decision was made to not update the framework and implementation guidelines prior to the commencement of the Sixth Assessment Report (AR6) cycle, which concludes in 2021/2022.

Despite the absence of a formal update to the framework, scholarship published in the wake of the last major assessment report (AR5) highlights how the framework was applied unevenly between working groups and chapter teams (Adler and Hirsch Hadorn 2014; Mach et al. 2017). Commentaries have also criticized the framework for lacking clarity about how authors are supposed to interpret the relationships between the framework's three language scales: evidence/agreement, confidence, and likelihood (Aven 2019; Aven and Renn 2015; Helgeson, Bradley, and Hill 2018; Winsberg 2018; Wüthrich 2017). However, the literature on the common challenges authors experience interpreting and applying the framework have, thus far, focused exclusively on the major assessment reports, with no consideration of the three special reports (SRs) published since AR5: Global Warming of 1.5°C (SR15) (IPCC 2018), Climate Change and Land (SRCCL) (IPCC 2019c), and The Ocean and Cryosphere in a Changing Climate (SROCC) (IPCC 2019b). These SRs constitute the latest application of the uncertainty language framework and perhaps provide the clearest indication of whether the issues raised since AR5 will be addressed in AR6.

While a significant amount of the literature on the IPCC uncertainty language framework has been written by scholars that have participated in the IPCC assessment process themselves (Adler and Hirsch Hadorn 2014), many of the criticisms are based on authors' personal experiences or anecdotal examples of uncertainty language drawn from reports. One exception is a systematic uncertainty term usage analysis of AR4 and AR5 conducted by Mach et al. (2017), which identifies trends in the application of the framework to empirically ground a series of recommendations for improving the framework. These trends include an increase in the number of uncertainty terms per report, the growing preference for confidence terms, and the decrease in “low certainty” statements in the Summary for Policymakers (SPM).

This article extends this term usage analysis to the three most recent SRs, confirming that many of these trends identified in AR4 and AR5 continue in the SRs. While the Mach et al. analysis focuses exclusively on SPMs and chapter executive summaries, the term usage analysis presented in this article also examines the application of the uncertainty language framework in the chapter bodies, revealing key differences in language choices between report chapters and the much more widely read SPMs. This finding suggests that the uncertainty language framework is applied differently in different parts of the reports by the same authors.

While the term usage analysis provides a picture of how the uncertainty language framework is being applied, I then present findings from interviews with IPCC coordinating lead authors (CLAs) and lead authors (LAs) on their interpretation of the framework and other decisions that underpin these trends. Author experiences confirm the claim made in the literature that inconsistencies in the usage of the confidence and likelihood scales stem from two fundamentally different interpretations of the framework (Wüthrich 2017). They also highlight how many of the idiosyncratic applications of the framework and inconsistencies between chapters and reports reflect factors such as time pressure, the absence of effective oversight, and the lack of emphasis placed on uncertainty language relative to other tasks.

Recent scholarship in the risk science literature advocates a more fundamental overhaul of the framework, criticizing the IPCC uncertainty language framework for possessing a somewhat convoluted interpretation of probability (Aven and Renn 2015; Aven 2019), mirroring the critique of the W&H framework that I present in Article #1. Therefore, the recommendations I propose to clarify and amend the IPCC uncertainty language framework take two tacks. First, I provide recommendations that carefully sidestep these epistemological concerns and analyze the application of the IPCC uncertainty language framework on its own terms. My primary aim here is to provide pragmatic recommendations to clarify the existing framework with the hope that they could be applied to AR6, as well as subsequent special reports that may be initiated before a new framework can be introduced. Second, I reflect on what a more consistent and rigorous framework might look like, given the conceptualization of uncertainty I present in Article #1.

This article is organized as follows. Section 4.4 summarizes the literature on the IPCC uncertainty language framework to date, focusing on recent scholarship addressing the implementation of the system during the production of AR5. Section 4.5 presents the results from the uncertainty term usage analysis for SR15, SRCCL, and SROCC, reinforcing a number of key questions about how the framework is interpreted and applied by IPCC authors. Section 4.6 seeks to answer these questions by analyzing the recent experiences of IPCC authors implementing the framework. Section 4.7 concludes by proposing several recommendations for clarifying and amending the framework to address persistent sources of confusion highlighted by the authors.

4.4 The IPCC uncertainty language framework

4.4.1 Evolution of the framework

Several articles have chronicled the history of the IPCC uncertainty language framework in detail (e.g. Swart et al. 2009; Mastrandrea and Mach 2011), including the initial development of the framework leading up to the third assessment report (TAR) (Moss and Schneider 2000), as well as subsequent refinements and clarifications for AR4 (IPCC 2005) and AR5 (Mastrandrea et al. 2010). The framework emerged from observations that chapter teams adopted a plurality of approaches to characterize and communicate uncertainty in the first and second assessment reports, leading to calls for a more systematic approach that could be applied across chapters and working groups. The most recent iteration of the framework provides three distinct but related scales that authors can use to qualify uncertain knowledge claims (Fig. 12).

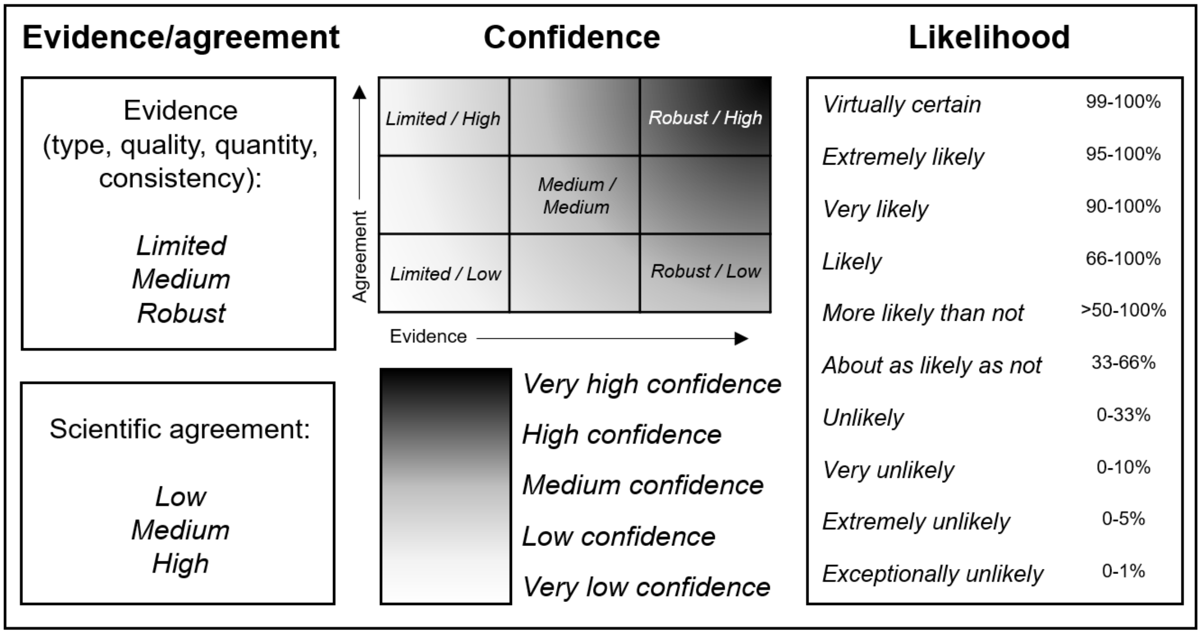

Figure 12: IPCC uncertainty language framework (reproduced from Mastrandrea et al. (2010) and Mach et al. 2017))

The *evidence/agreement* scale allows authors to separately assess the type, amount, and quality of the evidence base supporting a claim and the level of scientific agreement in the literature (both on three-point scales). The framework then specifies that evidence and agreement statements be presented together (e.g. “limited evidence, medium agreement”). The guidance note provides some suggestions for how to evaluate evidence and agreement – however, these assessments are highly dependent on the specific context of the topic area and research landscape – and are thus left largely to the expert judgement of the authors.

The five-point *confidence* scale is closely tied to the evidence/agreement scale. According to the guidance note: “Confidence in the validity of a finding [is] based on the type, amount, quality, and consistency of evidence and the degree of agreement” (Mastrandrea et al. 2010). According to this definition, confidence statements are derived from combining and integrating the evidence and agreement assessments. For example, a “limited evidence, low agreement” assessment might be translated as “low confidence.” However, as illustrated in the three-by-three matrix in Figure 12, there are a number of evidence/agreement combinations (like “medium evidence, high agreement”) where it is not entirely clear how to translate between the scales.

Lastly, the *likelihood* scale is used to communicate quantified, probabilistic assessments of uncertainty produced by statistical or modeling analyses, or formal expert elicitation methods. The scale provides 10 likelihood terms that are each attached to a specific probability interval (Fig. 12). In addition to presenting the three uncertainty language scales that make up the framework, the accompanying guidance (Mastrandrea et al. 2010, 2011; Mastrandrea and Mach 2011) also provides best practices for how to implement the framework. For example, authors are encouraged to use the confidence scale when there is robust evidence and high agreement (Mastrandrea and Mach 2011) – implying that authors should use the confidence scale sparingly when the evidence base is weaker or there is less agreement in the literature. Additionally, authors are encouraged to use the likelihood scale to express quantified uncertainty assessments when there also exists "sufficient confidence" – suggesting that all likelihood assessments have an implicit "high" or "very high" confidence assessment attached to them. Therefore, the guidelines hint at an implicit hierarchy between the scales (Swart et al. 2009).

Note that the terms "likely" and "likelihood" are used in Article #1 (and in most scholarship dealing with uncertainty and probability) to describe *any* situation characterized by uncertainty – whether uncertainty can be quantified or not. However, by calling this scale the "likelihood scale," the IPCC takes the rather unique approach of designating likelihood as a strictly quantitative concept that is measured and communicated probabilistically. The result of this decision is that it is fairly ambiguous whether or not confidence levels are also an expression of probability (which is the position taken in Article #1) or if confidence is somehow a qualitatively distinct (and inferior) expression of uncertainty according to the IPCC's framework.

Recent scholarship in the philosophy of science and risk science literature also argues that the IPCC's framework rests on a somewhat convoluted interpretation of probability (Aven and Renn 2015; Aven 2019).[25] The guidance note specifies that "confidence should not be interpreted probabilistically" (Mastrandrea et al. 2010, pp. 3). However, according to the subjective (i.e. Bayesian) interpretation of probability – which has emerged as the dominant paradigm in most scientific fields today (Morgan 2014), a decision maker's subjective assessment of their confidence (or degree of belief) in the truthfulness of a knowledge claim is, in fact, the very definition of probability. From this perspective, it is not clear why confidence statements cannot also be quantified and expressed probabilistically (Aven and Renn 2015). More fundamentally, as I argue in Article #1, the frequentist and subjectivist interpretations of probability

[25] For a more detailed discussion of Aven's (2019) interpretation of the confidence and likelihood scales, see Appendix B.

are based on incompatible theories of knowledge and any attempt to reconcile them is bound to create confusion.

From its inception, the uncertainty language framework was viewed as an iterative process subject to ongoing improvements, with the first guidance note stating: "guidelines such as these will never truly be completed" (Moss and Schneider 2000, pp. 34). The decision to not amend the framework between AR5 and AR6 was intentional, reflecting a desire to maintain consistency between the two reports.[26] While there have been no recent updates to the framework, commentary on the application of the framework in AR5 echoes many of the same critiques levelled at the application of the framework in AR4 – namely that the framework was applied inconsistently across chapter teams and working groups (Adler and Hirsch Hadorn 2014; Mach et al. 2017) and confusion persists around the three-scale framework itself (Aven and Renn 2015; Helgeson, Bradley, and Hill 2018; Wüthrich 2017; Borges de Amorim and Chaffe 2019). Other prominent critiques like the framework's poor treatment of the concepts of risk and surprise (Aven and Renn 2015; Aven 2019) and concern that the framework no longer reflects the evolving purpose of the IPCC assessment process (Beck and Mahony 2018; Oppenheimer et al. 2007; Kowarsch and Jabbour 2017) are beyond the scope of this paper.

4.4.2 The relationship between the evidence/agreement, confidence, and likelihood scales

According to the uncertainty framework guidance, confidence assessments are made by integrating the quantity and quality of the underlying evidence base with the level of agreement in the literature using the three-by-three matrix summarizing the relationship between the two scales (Fig. 12). However, the framework is far less clear about when and when not to translate evidence/agreement statements into confidence statements (Wüthrich 2017). Also, if specific combinations of evidence/agreement terms directly translate into confidence terms, it is not apparent why there needs to be two separate scales at all.

In a follow-up article to the most recent uncertainty language guidance note, two of the framework's authors offer a further explanation of how to deploy the two scales together, which appears to justify the existence of two scales that communicate similar uncertainty information. They suggest that confidence language should be used "[f]or findings associated with high agreement and much evidence or when otherwise appropriate" (Mastrandrea and Mach 2011, pp. 663). The guidance note itself specifies that "the presentation of findings with "low" and "very low" confidence should be reserved for areas of major concern" (Mastrandrea et al. 2010, pp. 3) Combining these two guidelines, it appears that stronger

[26] Source: interviews with members of the IPCC Bureau.

evidence/agreement terms should always be translated into confidence language, while weaker evidence/agreement terms should only be translated when the claim is particularly salient to policy makers. However, the terms "robust evidence, high agreement" and "low confidence" frequently appear in the chapters of IPCC reports – a practice that is shaped by "somewhat arbitrary working-group and disciplinary preferences" (Mach et al. 2017, pp. 9).

Meanwhile, the relationship between the confidence and likelihood scales has received considerably more attention in the literature (Committee to Review the IPCC 2010; Curry 2011; Jonassen and Pielke 2011; Jones 2011; Mastrandrea and Mach 2011; Mach et al. 2017; Helgeson et al. 2018). The nature of the confusion around how to distinguish between the confidence and likelihood scales and how they fit together has been most effectively described by Wüthrich (2017) who identifies two dominant interpretations of the relationship in IPCC assessment reports, which he labels the *substitutional* and *non-substitutional* interpretations.

The substitutional interpretation sees both the confidence and likelihood scales as providing the same basic information – that is, a measurement of the validity or "truthfulness" of the knowledge claim or finding. Translated into Bayesian terms, both scales communicate the assessor's degree of belief that the claim accurately reflects the "real world." From this perspective, an author uses likelihood language when uncertainty has been formally quantified by a statistical analysis, model, or expert elicitation – and qualitative confidence language when there is a lack of quantitative evidence. But the two scales are essentially substitutable, despite the fact that the likelihood scale is generally used to describe frequentist probabilities and the confidence scale is used to describe subjective probabilities. A "very high confidence" statement can be considered to be equivalent to a likelihood statement of "extremely likely" or "very likely," with the only difference between the two scales being the nature of the evidence informing the assessment. This perspective reflects the version of the uncertainty language framework used in AR4, which attached quantitative indicators to the qualitative confidence labels like "about 8 out of 10 chance" (IPCC 2005). From this perspective, an author might address the issue of model unreliability or structural uncertainty by "downgrading" a likelihood term with a more precise probability interval (e.g. "very likely") to a term with a less precise interval (e.g. "likely").

The non-substitutional interpretation, which Wüthrich suggests is used in the majority of cases, only sees confidence as an assessment of the validity or truthfulness of a finding. From this perspective, likelihood language should only be used to refer to an outcome from a specific statistical study or model (or ensemble of models) that produces probabilities. For example, the statement "Between 1979 and

2018, Arctic sea ice extent has *very likely* decreased for all months of the year" (IPCC 2019a, pp. 4) simply means that a specific study showed that Arctic sea ice has decreased in 90-100% of observations or model runs. The finding from the study is but one piece of evidence which may (or may not) influence an author's assessment of the validity of the claim (i.e. their confidence assessment). It is conceivable that the model was poorly constructed or there is high structural uncertainty and as a result, an assessor will have low confidence in the finding despite the fact that the study assigned a high probability value to the outcome. Therefore, according to the non-substitutional interpretation, a likelihood assessment is not an assessment at all but the presentation of a specific probabilistic finding – or using Wüthrich's language, confidence statements are "meta-judgements" while likelihood statements are "intra-finding judgements."

Winsberg (2018) proposes a similar interpretation of the confidence/likelihood distinction, suggesting that likelihood terms be understood as "first-order probabilistic claims" (based on statistical and modeling analyses), while confidence terms are "second-order probabilistic claims" that describe the "resilience" of the likelihood assessment (i.e. the quality of the statistical and modeling analyses) – despite the fact that the IPCC uncertainty language guidance specifically cautions authors to not interpret confidence statements probabilistically.

According to implementation guidance written by several authors of the framework (Mastrandrea et al. 2011), likelihood statements should be used when confidence is high or very high, which has been interpreted by some commentators that lone likelihood statements contain implicit (high) confidence assessments in the quality of the model (Helgeson, Bradley, and Hill 2018; Mastrandrea and Mach 2011). The implication is that authors are discouraged from communicating findings from modeling studies when they have medium or low confidence in the quality of the model. If this interpretation is correct, than we can assume that when confidence and likelihood statements are paired together, the confidence statement refers to the knowledge claim, and the author's high confidence in the quality of the model is implicit.

Consider, again, the statement about the observed decrease in Arctic sea ice. Adopting the substitutional interpretation, the assessment of the validity of the finding is straightforward: the term "very likely" is equivalent to having "very high confidence" that Arctic sea ice has decreased all months of the year. Therefore, the term "very likely" is both the output of a model, as well as the author's assessment of the validity of the finding.

But from the non-substitutional interpretation, "very likely" simply means that one model generated the probability interval 90-100%. And therefore, the author is still responsible for assessing the validity of that claim using confidence language. One might argue that if an author is highlighting a particular study or model that makes such a bold claim, then clearly they should have high confidence in not only the quality of the model but in the finding itself. The implication is that all likelihood statements carry not one but two implicit (high) confidence assessments. However, this relationship is far less clear with the likelihood statement "more likely than not" that describes a much wider probability interval (50-100%). It seems far less obvious that an author would have high confidence in a finding supported by such an imprecise probability interval – even if they had high confidence in the quality of the model.

The main takeaway from these critiques is that there continues to be confusion over the relationships between the three scales. So far, the literature highlighting these issues has generally appealed to anecdotal examples of how the scales have been used in IPCC reports. However, a more systematic, empirical analysis of uncertainty language usage – like the study conducted by Mach et al. (2017) – provides a much firmer basis for recommendations for improving the framework. The following section presents an analysis of uncertainty term usage in the three most recent SRs, highlighting the key trends in language usage since AR4.

4.5 Uncertainty term usage in IPCC special reports

This section analyzes the use of the designated uncertainty terms in the SR15, SRCCL, and SROCC,[27] building off a study conducted by Mach et al. (2017) comparing uncertainty language usage in AR4 and AR5. The language usage trends discussed in this section are investigated further in Section 4.6, which presents findings from interviews with IPCC authors on their experiences interpreting and applying the framework in recent SRs.

4.5.1 Methods of analysis

Major assessment reports like AR4 and AR5 are composed of three sub-reports produced by the IPCC's three working groups tasked with addressing different components of the climate change issue: the physical science basis (WGI), impacts, adaptation, and vulnerability (WGII), and mitigation (WGIII). Working group sub-reports are made up of approximately 15 to 30 chapters produced by separate author teams, as well as an SPM that presents the key findings from all chapters. By comparison, the SRs contain

[27] The analysis preceded the publication of the final drafts of the SRCCL and SROCC and is based on the "approved drafts," which were subject to final edits and tricklebacks. Trickleback documents for the SRCCL and SROCC contain only 25-30 suggested revisions associated with uncertainty language each. Therefore, the term usage in the final drafts may vary slightly from the data used in this analysis but will not significantly affect the findings presented here.

no sub-reports and include a mixture of new authors and authors that previously contributed to WGI, WGII, and WGIII. The SRs are significantly shorter than the major assessments, ranging from just five to seven chapters in length (plus an SPM).

In addition to comparing uncertainty language usage in the SPMs and chapter executive summaries, I also analyzed the uncertainty statements made in chapter bodies. Incorporating uncertainty statements from the chapter bodies significantly increases the total number of uncertainty statements in the analysis and provides insight into the decision-making of author teams as they determine which knowledge claims (and associated uncertainty statements) to "elevate" to the executive summaries and SPM, which are read by a much larger audience.

For each chapter body, chapter executive summary, and SPM, I tabulated the number of unique instances of evidence/agreement, confidence, and likelihood terms. I analyzed specific terms according to the rate at which they were used (i.e. number of unique instances per page) and according to their proportion of the total number of uncertainty terms used in each chapter and report.

I also analyzed the frequency that authors used confidence and evidence/agreement terms of various "certainty levels." Following the criteria outlined by Mach et al. (2017), terms were grouped into three categories: low certainty, medium certainty, and high certainty. Lastly, I analyzed working group participation of SR CLAs and LAs in major assessment reports using the IPCC author database (IPCC 2019a), which includes author information for AR4, AR5, and AR6.

4.5.2 Results and discussion of uncertainty language trends in IPCC special reports

The overall number of uncertainty terms appearing in SPMs has steadily increased from AR4 (an average of 4.3 terms per page) to the two most recent SRs published in 2019 (11.0 and 10.8 terms per page in the SRCCL and SROCC). Figure 13a shows the proportional usage of the evidence/agreement, confidence, and likelihood scales for the SPMs of the three SRs and compares them to the SPM of each working group for AR4 and AR5 based on data from Mach et al. (2017). Much of the analysis conducted by Mach et al. compares the language use of the three working groups. However, SR15, SRCCL, and SROCC all have a mixture of CLAs and LAs with previous WGI, WGII, and WGII experience.[28] While a significant proportion of the CLAs and LAs from each SR did not participate in AR4, AR5, or AR6 (SR15: 23%;

[28] The IPCC author database does not provide information on author participation in the FAR, SAR, or TAR.

SRCCL: 51%; SROCC: 61%), WGII is the most represented working group for all three SRs, followed by WGIII, and then WGI (Fig. 13c).

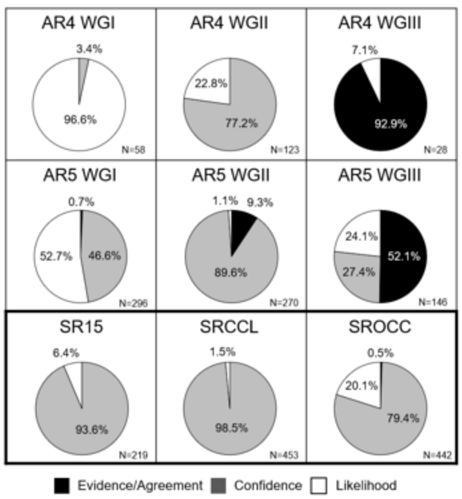

Fig. 13a: comparison of proportional usage of uncertainty language scales in SPMs (data for AR4 and AR5 from Mach et al. 2017)

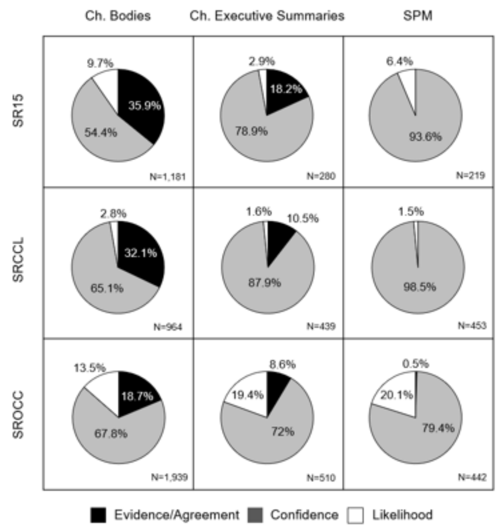

Fig. 13b: comparison of proportional usage of uncertainty language scales in chapter bodies, executive summaries, and SPMs

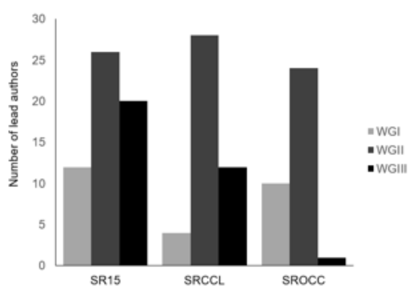

Fig. 13c: AR4, AR5, and AR6 working group participation by lead authors of special reports

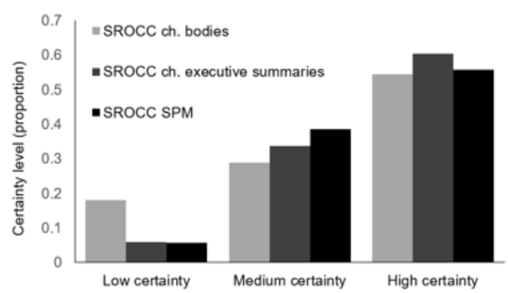

Fig. 13d: proportion of low, medium, and high certainty terms in the SROCC

Figure 13: Uncertainty term usage in recent IPCC special reports

The combination of working groups in the SRs makes it difficult to track some of the observations made by Mach et al., such as WGI's higher use of likelihood terms than WGII and WGIII in both AR4 and AR5. One might expect that SR15 – which has the most WGI authors of the three SRs – would also contain a significant number of likelihood statements. However, likelihood statements make up a mere 6.4% of all uncertainty statements in the SR15 SPM. But upon closer inspection of the underlying chapter bodies, 71% of all likelihood language in the SR15 is found in Chapter 3, which has nearly half (42%) of all the WGI authors in the entire report, which is consistent with the observed higher rate of usage of likelihood language by WGI authors in AR4 and AR5.

The proportional usage of the three uncertainty language scales in the SPMs of SR15, SRCCL, and SROCC (Fig. 13a) is consistent with the observed increase in the proportional use of confidence language across all working groups from AR4 to AR5. The proportional term usage profiles for the SPMs of the three SRs are also quite similar to one another, with confidence language accounting for 79.4%, 93.6%, and 98.5% of uncertainty language in the SR15, SRCCL, and SROCC SPMs respectively. The proportional term usage in the SRs also closely resembles the term usage in the WGII SPMs for both AR4 and AR5, which should not be surprising considering the large WGII representation in all three SRs (Fig. 13c).

While the use of evidence/agreement statements in AR4 and AR5 SPMs fluctuates considerably between reports and working groups, it has almost disappeared in the SR SPMs. Not a single evidence/agreement statement appears in the SR15 SPM, while evidence/agreement terms make up just 1.5% and 0.5% of all uncertainty terms in the SRCCL and SROCC SPMs respectively. However, the use of the evidence/agreement scale is still prevalent in the underlying chapter bodies of the SRs. In fact, one can observe a regressive use of the evidence/agreement scales moving from the chapter bodies to executive summaries to the SPM (Fig. 13b). The same pattern can be observed with “low certainty” evidence/agreement and confidence terms, which are fairly prevalent in chapter bodies but nearly disappear in executive summaries and SPMs. This trend is illustrated for the SROCC in Figure 13d.

The term usage analysis also reveals variation in the application of the uncertainty language framework between chapters within the same report. Figures 14a, 14b, and 14c show the proportional usage of the evidence/agreement, confidence, and likelihood scales for each chapter of the SR15, SRCCL, and SROCC, as well as the per page usage of uncertainty terms in the chapter bodies. One chapter in SR15[29] and one chapter in SROCC[30] contain approximately twice as many uncertainty terms per page than the next most language-dense chapter. The common thread between these two chapters is the extensive use of figures and tables to communicate large amounts of evidence/agreement and confidence statements that are not replicated in the chapter text. The use of colors, shadings, and symbols in figures and tables to convey uncertainty statements appears to be a relatively recent innovation in IPCC assessment reports.

[29] SR15 Chapter 5: Sustainable Development, Poverty Eradication and Reducing Inequalities
[30] SROCC Chapter 5: Changing Ocean, Marine Ecosystems, and Dependent Communities

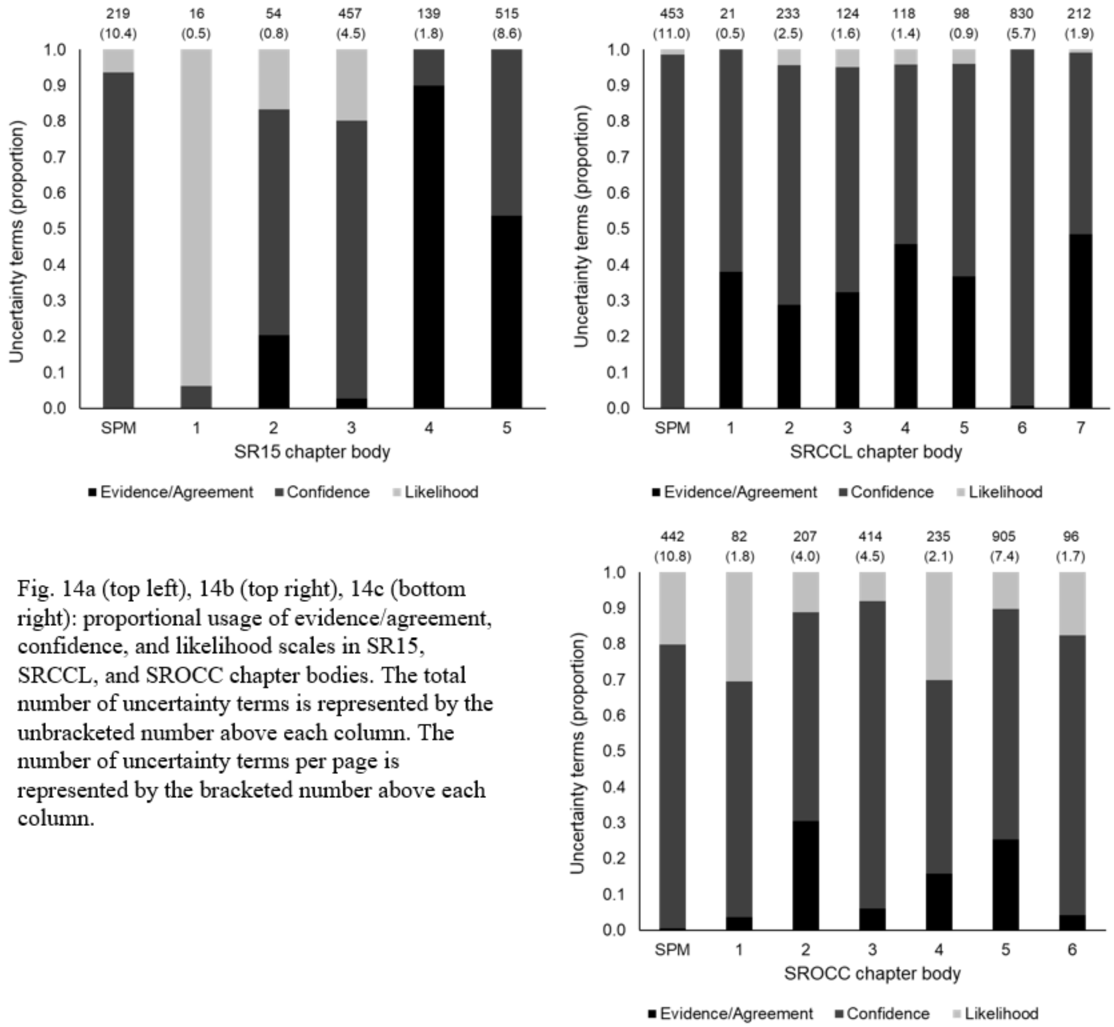

Fig. 14a (top left), 14b (top right), 14c (bottom right): proportional usage of evidence/agreement, confidence, and likelihood scales in SR15, SRCCL, and SROCC chapter bodies. The total number of uncertainty terms is represented by the unbracketed number above each column. The number of uncertainty terms per page is represented by the bracketed number above each column.

Figure 14: Proportional usage of uncertainty terms in SR15, SRCCL, and SROCC

Looking beyond the chapter executive summaries to their underlying chapter bodies reveals many unusual applications of the uncertainty language framework by specific chapter teams. One potentially problematic example of idiosyncratic language use is the unintentional use of unbracketed and unitalicized likelihood language. For example, Chapter 5 of the SRCCL states: "… large shifts in land-use patterns and crop choice will likely be necessary to sustain production growth and keep pace with current trajectories of demand" (IPCC 2019b, pp. 5-28). Here, the term "likely" does not seem to be attached to a particular probability interval or finding from a modeling study, which could confuse an audience that is instructed to interpret likelihood terms as assessments of quantitative evidence.

Lastly, there are also many examples where uncertainty statements are combined or integrated when moving from the chapter body to the executive summary. In these situations, two distinct but related knowledge claims presented in a chapter body – each with their own uncertainty term – appear to be integrated into a single knowledge claim with a single uncertainty term in the executive summary. The integration of two specific claims into a broader knowledge claim can be considered a "meta-assessment," whereby the assessor assigns an uncertainty term (typically a confidence term) that applies to the aggregated evidence and agreement of the two sub-claims. This practice is common across many SR chapters and reveals an unexplored dimension of the IPCC assessment process.

4.5.3 Conclusions

The term usage analysis of SR15, SRCCL, and SROCC reinforces a number of trends seen in AR4 and AR5, including the overall increase in uncertainty terms per report, working group-specific preferences for particular uncertainty language scales, and the emergence of confidence language as the dominant scale for communicating uncertainty in IPCC assessment reports. Additionally, it reveals a number of other interesting findings that could indicate how the framework is being applied by IPCC authors in the preparation of AR6. These findings include: the disappearance of evidence/agreement statements in SPMs (despite their continued use in chapter bodies and executive summaries), the use of color, shading and symbols in figures and tables by particular chapter teams to communicate large amounts of uncertainty statements, and the meta-assessment of multiple uncertainty statements as findings are passed up from chapter bodies to executive summaries and the SPM.

While the term usage analysis points to several interesting trends, it does not provide much insight into the decisions, factors, and interpretations underlying them. The following section builds off this analysis by describing the recent experiences of IPCC authors interpreting and applying the uncertainty language framework.

4.6 Author experiences using the uncertainty language framework

This section explores the experiences of IPCC authors applying the uncertainty language framework during the production of the two most recent SRs: the SRCCL and SROCC. The discussion is informed by a series of semi-structured interviews conducted with CLAs and LAs from chapters 4 and 6 of the

SRCCL[31] and chapters 2 and 5 of the SROCC,[32] as well as members of the IPCC Bureau responsible for training and supporting authors on the implementation of the framework (N=14).[33] Interviewees were asked to reflect on the language usage trends identified in Section 4.5, why the framework is applied inconsistently within and across chapter teams, and how they interpret the relationships between the three language scales.

The interviews were conducted following the approval of each report's SPM but before trickleback revisions[34] and final copy-edits were applied to individual chapters. Chapter 6 of the SRCCL and chapter 5 of the SROCC were both selected because of their extensive use of colour-coded tables to communicate large amounts of calibrated uncertainty language. The other two chapters were selected because they reflect the "typical" or "average" chapter in each report in terms of their use of uncertainty language.

4.6.1 Decisions and dynamics underpinning language usage trends

Interviewees were unsurprised by the observed trend towards greater use of the confidence scale in recent IPCC assessment reports. They attributed the decrease in likelihood language in the SRCCL and SROCC to the nature of the relevant evidence base for most of the topics covered by the two reports and the relative lack of WGI authors. For instance, one author stated:

> WGI tends to use a lot of likelihood statements because they lean on models producing numerical outputs. But I work in the mitigation space where a lot of what we're looking at is not numerical – we're looking at things like sustainability, social impacts, and gender influences where it is more appropriate to use levels of evidence and agreement, or to boil those down into confidence statements.

However, authors tended to attribute the decline in evidence/agreement language (and the related increase in confidence language) in the SPM to the tendency of chapter teams to elevate a greater proportion of higher certainty claims into the executive summary and SPM. This explanation would suggest that authors are following the recommendation in the guidance to translate claims with strong evidence and agreement into confidence language. The significantly higher proportion of medium and

[31] SRCCL Chapter 4: Land Degradation; SRCCL Chapter 6: Interlinkages between Desertification, Land Degradation, Food Security and GHG fluxes: synergies, trade-offs and Integrated Response Options
[32] SROCC Chapter 2: High Mountain Areas; SROCC Chapter 5: Changing Ocean, Marine Ecosystems, and Dependent Communities
[33] Interviews by chapter/group: SRCCL Ch. 4 (4); SRCCL Ch. 6 (2); SROCC Ch. 2 (3); SROCC Ch. 5 (3); IPCC Bureau (2). Interview requests were sent to all CLAs and LAs for each chapter and were conducted with all individuals that responded.
[34] A series of final revisions of report chapters following the SPM approval meetings.

high certainty terms in the SPM and chapter executive summaries compared with the underlying chapter bodies (Fig. 13d) supports this explanation as well. While authors were not expressly discouraged from using evidence/agreement or "low confidence" qualifiers in the SPM, all interviewees supported the notion that SPMs should largely highlight areas where scientific knowledge is most robust, rather than highlighting knowledge gaps.

Commenting on how SPMs rarely highlight low certainty claims, one interviewee described their thought process for determining which findings should be proposed for the SPM, saying: "We often perform a reality check and say: 'if this is really a low confidence issue, why are we having it in the SPM?' – because space is really at such a premium." Another author described how demand-side pressure for higher certainty claims from policy makers also plays a role in the disproportionally high number of high certainty (and confidence) statements in the SPM:

> We only want the most robust messages in the SPM because we have to go through government approval – and if we have a whole bunch of low confidence statements in there, the governments can rightly ask, "if we have low confidence in this, why are you telling us about it? Tell us the things you know, not the things you don't know."

When asked whether it is important for scientific assessments to also present important knowledge gaps, authors tend to point towards the IPCC's stated purpose – to support policy makers – suggesting that pressing policy decisions are better supported with high certainty claims than identified knowledge gaps. However, there was also recognition that in certain situations, low certainty statements may be highlighted in the SPM. These situations tend to involve claims that are of particular interest to policy makers and their inclusion must be vigorously defended by the author.

The observation in the language usage analysis that author teams appear to be conducting meta-assessments, where multiple claims from the chapter body are combined into a broader claim with a single uncertainty statement in the executive summary or SPM, was confirmed by interviewees as a "common practice." One author noted that these meta-assessments were particularly important when combining knowledge claims from multiple chapters in the SPM and that their team was very careful to maintain a transparent "line of sight" between the meta-assessment and the underlying evidence in the chapter body.

Large colour- and symbol-coded tables and figures are used in both chapter 6 of the SRCCL and chapter 5 of the SROCC to communicate a large number of knowledge claims and their associated uncertainty assessments. Tables illustrating the interactions between various systems and variables use colours, shadings, and symbols to communicate the strength or direction of those interactions, as well as assessments of the uncertainty associated with each relationship. Authors of these chapters explained that the primary purpose of using tables (as opposed to merely describing each relationship in the text) is to communicate the comparative nature of the analysis. However, authors contributing to these chapters also expressed a desire for more guidance on visual communication. While an external design firm is brought in to improve and harmonize figures, diagrams, and tables in the SPM, an interviewee commented that the authors are mostly "winging it" in the chapter bodies.

The unintentional or colloquial use of likelihood terms was also identified as a common problem that is surprisingly difficult to recognize and avoid:

> There were several times in text drafting where my natural preference was to summarize the balance of evidence as either "unlikely" or "likely." However, it would have been inappropriate for me to do so, since these terms need to be used in a precise, quantitative way. Thus, rewording had to be along the lines of "evidence is unconvincing that…" or "it is expected that…"

Other examples of idiosyncratic language use such as lone evidence or agreement terms were dismissed by interviewees as reflecting either "sloppiness" stemming from time constraints or "a lack of proper training." Interviewees appealed to the significant time constraints built into the IPCC assessment process more than any other factor for explaining the causes of inconsistent language usage. The vast majority of authors conduct IPCC-related work above and beyond their normal professional duties as academics, researchers, or government scientists and spend hundreds of hours writing, reviewing comments, and approving text on evenings and weekends. These time constraints are magnified for SRs, which are prepared on much shorter timelines than the major ARs, with one interviewee suggesting that the most recent SRs "do not necessarily reflect the latest implementation of the IPCC uncertainty guidelines but the implementation of the guidelines under pressure."

A number of interviewees also pointed to the training process as an area that might be improved going forward. Several authors that had participated in previous IPCC major ARs commented that the presentations delivered at each of the first three lead author meetings for the SRCCL and SROCC were a significant improvement over previous assessment cycles. However, one suggestion echoed by multiple interviewees was to conduct a series of chapter-specific practice examples with each author. Other suggestions focused on the support and oversight that authors receive while developing the chapter text. One author suggested that it would be helpful to have better editing or "policing" in the final stages of revision from an individual outside of the chapter team – but acknowledged the already immense workload of co-Chairs. Another interviewee suggested that perhaps the chapter's review editors – whose main responsibility is ensuring that authors address all submitted comments – might also play a bigger role in monitoring the implementation of the uncertainty framework.

A final factor that was pointed to as an underlying cause of the inconsistent or "sloppy" application of the framework was the lack of emphasis that uncertainty language received relative to more pressing tasks like meeting deadlines and ensuring that chapter sections comprehensively address the literature. This sentiment was particularly evident in comments about how uncertainty qualifications were sometimes left to the very end of the writing process, as well as comments about how deliberations at SPM approval meetings rarely focus on specific uncertainty terms.

4.6.2 **Translating between the three scales**

4.6.2.1 When to use uncertainty language

Nearly all interviewees expressed support for the uncertainty language framework as a reasonably effective system for characterizing and communicating uncertainty.[35] However, several authors commented that while the framework is strong "on paper," a number of issues arise during its implementation. One example is the issue of determining when a knowledge claim requires an uncertainty statement versus merely providing citations. Nearly every sentence in an IPCC assessment report relates to the behaviour of physical and social systems. Therefore, authors must decide if each sentence is merely a statement of widely agreed-upon fact (requiring no supporting evidence or uncertainty statement), a policy-relevant knowledge claim (requiring an uncertainty statement), or a sub-claim made in the literature that supports the broader policy-relevant claim (requiring the citation of supporting evidence).

Interviewees describe these decisions as highly subjective. For example, one author asked: "Where is the boundary between 'established fact' and 'very high confidence'?" A few authors alluded to a hierarchy of claims, where more specific claims made in the literature, such as particular numbers, were given citations and more general claims that were supported by such numbers were qualified with uncertainty language. One interviewee explained how decisions about when to apply uncertainty language may depend on the geographical scale of the claim:

> For a global phenomenon like global sea level rise, it is extremely important that you have uncertainty language so you can document

[35] A sentiment that is not echoed by many risk/uncertainty scientists (Aven 2019).

> exactly how and why you arrived at that conclusion. But as you make it a finer and finer resolution issue, it becomes a fractal problem – the question about uncertainty for specific ecosystems should often be answered with "it depends."

Others appealed to "rules of thumb" to help guide their decision-making about which sentences to assign uncertainty qualifiers. One author said they had received strong encouragement from the chapter's CLA to conclude each section with a sentence that has uncertainty language in it. Others described how their chapter teams decided to omit uncertainty statements and then retrofit their sections with uncertainty language once they had a better idea of which findings would be pushed up into the chapter's executive summary and would apply uncertainty language to these higher visibility claims.

4.6.2.2 Evidence/agreement vs. confidence

Author testimony confirms the claim made throughout the literature that the distinctions between the three language scales are interpreted differently by different chapter teams, and even by different authors within the same chapter team. As discussed in Section 4.4.2, the framework is somewhat vague about when to use evidence/agreement statements and when to translate them into confidence statements. Author responses to questions about when and why they deployed the evidence/agreement scale versus the confidence scale reveal two main interpretations of the relationship between the two scales.

The first interpretation, which was shared by the majority of interviewees, sees the relationship as more or less formulaic in nature. According to the formulaic interpretation, every evidence/agreement statement has a "proper" corresponding confidence statement as defined by the three-by-three matrix in the guidance note (Fig. 12). For uncertainty statements with matching levels of evidence and agreement like "medium evidence, medium agreement," the conversion is straightforward ("medium confidence"), while a statement like "limited evidence, high agreement," can be converted using the matrix to "medium confidence." However, some judgement must be exercised for ambiguous terms like "robust evidence, medium agreement."

Authors expressing the formulaic interpretation employed one of two strategies to determine which scale to use while writing the report. First, a number of authors chose to translate all non-ambiguous evidence/agreement statements into confidence statements and only use the evidence/agreement scale for more ambiguous combinations like "robust evidence, medium agreement." The second strategy is to use evidence/agreement statements in the chapter body and only translate them into confidence statements in the executive summary and SPM. For example, this was the predominant strategy used by authors of chapter 4 of the SRCCL because it provided "a more explicit and differentiated assessment."

Meanwhile, the second interpretation of the relationship between the evidence/agreement and confidence scales sees a subtle qualitative distinction between the scales. While the two interviewees that shared this interpretation acknowledged that confidence assessments are based primarily on the strength of evidence and agreement, they also appealed to an additional dimension that is present in confidence assessments but not evidence/agreement assessments:

> The evidence/agreement scale is a comment about the literature, where you say: "don't come to me if it turns out wrong – but I'm just saying there's a lot of literature and it looks pretty good to me." Whereas confidence language is upping the ante – it's "owning" the assessment. There is an extra layer there. It's not just an algorithmic translation.

According to this explanation, the non-formulaic interpretation sees the confidence scale as expressing a degree of personal ownership over the assessment, while evidence/agreement statements allow authors to maintain a level of detachment from the assessment. From this perspective, the translation matrix is merely a guide. The non-formulaic interpretation seems to support the suggestion that evidence/agreement language should be used for less robust claims and confidence language for claims with strong evidence and agreement, while the formulaic interpretation sees these scales as essentially substitutable at the authors' discretion.

4.6.2.3 Confidence vs. likelihood

Interviewees were asked to explain the distinction between the confidence and likelihood scales and describe the circumstances when it is most appropriate to use each scale. Responses largely conformed to the non-substitutional interpretation described by Wüthrich (2017). However, one author's response perfectly articulated the substitutional interpretation that sees confidence and likelihood terms as interchangeable: "The confidence assessment is qualitative and the likelihood assessment is quantitative. But in practical terms, the descriptors "very confident" and "very likely" represent similar levels of trust that the findings provide an accurate reflection of the real world."

Meanwhile, interviewees that reflected the non-substitutional interpretation acknowledged the fallibility of models producing the probabilistic outputs on which likelihood statements are based – and the necessity of attaching an implicit or explicit confidence assessment to model outputs. For example:

> Model-derived, mathematically exact values are not quite what they seem, since they necessarily depend on the many assumptions and simplifications made in constructing the model. Thus, a "spurious precision" is involved, with a hidden, secondary level of confidence associated with the validity of those assumptions.

When asked whether likelihood statements should also be accompanied by a confidence statement about the quality of the underlying model or the validity of the finding, most authors agreed that the confidence/likelihood distinction was confusing but disagreed on how best to communicate probabilistic findings. Some interviewees supported the use of both scales to qualify individual statements, while others hinted that likelihood statements already contain an implicit "high confidence" assessment. One author felt strongly that attaching both a confidence and likelihood statement to the same claim – while possibly improving transparency and rigor – would ultimately be counter-productive: "Combining the two scales is going to be mostly confusing – both for the people having to come up with the statement and the people that have to interpret it." Meanwhile, an author contributing to chapter 2 of the SROCC described how their team navigated this problem by avoiding the use of likelihood language when confidence about model quality was not high, electing to use the confidence scale instead:

> For the glacier projections, we used a homogeneous data set you can analyze quantitatively. But we downgraded it to "medium confidence" because the models used for the 100-year projections are relatively simple. But we have not done that consistently and it wasn't clear whether we should do that all the time.

While the IPCC uncertainty guidance stresses both consistency and flexibility in how the framework is applied across chapters and reports, the clear message from authors is that they want more direction on how to interpret the framework.

4.6.3 Limitations of the study

While the CLAs and LAs interviewed in this study were drawn from multiple chapter teams from two different special reports with the goal of representing potentially diverse experiences and interpretations of the framework, the sample noticeably lacks WGI representation. Of the 14 interviews conducted, only two interviewees had previous experience contributing to a WGI report, which reflects the low participation of WGI authors in the SRCCL and SROCC overall. Both the language usage analysis and reflections of interviewees suggest that WGI authors tend to assess more quantitative studies and use likelihood terms more frequently than the other two working groups. Therefore, it is possible that WGI authors may interpret the distinction between the confidence and likelihood scales differently, for example. However, AR6 will present an opportunity to conduct a more precise comparison of how the uncertainty language framework is interpreted and applied by the three working groups.

The other potential limitation of the study is the generalizability of the experiences of IPCC authors working on special reports to the much longer and working group-specific assessment reports that make up the AR6. While many of the interviewees' responses reflect both their experiences working on the SRCCL or SROCC and their previous experience working on ARs, it is possible that the heightened time constraints associated with the SRs created conditions for interpreting and applying the framework that were unique to the SR process. At the same time, interviewees that had participated in previous ARs claimed that the training they received for the SRs was more extensive than training they had received previously. Therefore, I believe that the experiences of authors applying the uncertainty framework in recent SRs are largely indicative of the experiences of authors currently engaged in the AR6.

4.7 Improving the characterization and communication of uncertainty in IPCC assessment reports

The experiences of IPCC authors preparing the SRCCL and SROCC illuminate how author interpretations, attitudes, and decisions contribute to the observed trends in uncertainty language usage. Authors see the primary purpose of IPCC assessments as the communication of robust, high certainty findings, which has led to a growing preference for confidence language in chapter executive summaries and the SPM, and a corresponding decrease in evidence/agreement terms. One potential negative consequence stemming from the disappearance of evidence/agreement terms in the SPM is that policy makers may become decreasingly familiar with the evidence/agreement terms, leading to confusion when they consult the underlying chapters where the evidence/agreement scale is used far more frequently.

Meanwhile, author testimony also shows how decisions about when to employ uncertainty language is unavoidably subjective. While there is no perfect set of rules for deciding when to qualify a knowledge claim with an uncertainty statement as opposed to a citation, the practice of adding uncertainty terms after most of a section's text has already been written should be discouraged since this practice is inconsistent with the spirit of scientific assessment. As many authors noted, an assessment is not a literature review. The evaluation of the validity of knowledge claims is the core output of IPCC assessment reports and therefore, authors should start making such judgements from the very beginning of the process – even if uncertainty terms must be continuously revised as more evidence is assessed.

Authors continue to be confused about how the three uncertainty language scales fit together. The formulaic interpretation of the relationship between the evidence/agreement and confidence scales, which was shared by the majority of interviewees, sees the three-by-three matrix in the guidance note as the authoritative tool for translating between the two scales. The less popular non-formulaic interpretation asserts that confidence statements add an extra ingredient of "ownership" to the assessment above and beyond simply aggregating evidence/agreement statements. However, this perspective overlooks the fact that evidence/agreement assessments also involve subjective judgement. When making evidence/agreement assessments, authors must determine what constitutes "enough" research, "high quality" research, and "well agreed-upon" research – all of which implicate the author in the resulting statement of evidence and agreement. Therefore, ownership over the judgement is, in fact, present in both scales.

While most authors see the relationship between the confidence and likelihood scales as non-substitutional, where likelihood statements describe (largely frequentist) probabilistic findings from specific models or statistical analyses, there continues to be disagreement about whether a sole likelihood assessment conveys sufficient implicit information about the author's judgement or if an accompanying confidence statement is necessary.

A number of proposals have been made since AR5 to amend the IPCC uncertainty language framework to address this confusion. For example, Mach et al. (2017) suggest eliminating the confidence scale entirely and simplifying the evidence/agreement scale into five levels of "scientific understanding." Meanwhile, Helgeson et al. (2018) propose to eliminate confusion and logical inconsistencies between the confidence and likelihood scales by attaching confidence assessments specifically to the probability intervals (where less precise probability intervals are given higher confidence statements and more precise intervals are given lower confidence statements).

However, both of these recommendations involve significant changes to the current system that would need to be applied after the conclusion of the AR6 cycle. First, I provide pragmatic recommendations for clarifying the current framework with the hope that they can be applied in the final stages of the AR6 cycle, as well as subsequent special reports that may be initiated before a new framework can be introduced. Then, I propose a series of recommendations that address structural and institutional obstacles impeding the application of the framework. I conclude by reflecting on what a more consistent and rigorous uncertainty language framework might look like based on the conceptualization of uncertainty I present in Article #1.

4.7.1 **Clarifying the existing uncertainty language guidance for AR6**

The non-substitutional interpretation is the most common interpretation of the distinction between the confidence and likelihood scales by IPCC lead authors (Wüthrich 2017). Therefore, in the short-term, the uncertainty language guidance provided to IPCC authors should explicitly endorse this interpretation by providing the following three clarifications:

(1) Clarify the relationship between likelihood statements and the specific statistical/modeling analyses from which they are derived. If likelihood statements describe findings produced by particular models or statistical analyses and do not constitute assessments of those findings, then the uncertainty language guidance should encourage authors to clearly connect probability intervals to the models that produced them. For instance, the statement "According to the single longest and most extensive dataset, the LSAT increase between the preindustrial period and present day was 1.52°C (the *very likely* range of 1.39°C to 1.66°C)" (SROCC, ch. 2, pp. 3) clearly links the probability interval "very likely" (90-100%) to a specific statistical study.

(2) Clarify the relationship between likelihood statements and the underlying assessment of the quality of the specific statistical/modeling analyses. It should be made explicit in both the uncertainty language guidance given to authors as well as the SPM that likelihood statements always contain an implicit assessment of "high confidence" in the quality of the statistical or modeling analysis. Findings stemming from probabilistic analyses where the author has medium or low confidence in the analysis should be qualified using the confidence or evidence/agreement scales only.

(3) Clarify the relationship between likelihood statements and the overarching assessment of the validity of the finding. It should not be assumed that lone likelihood statements convey an implicit assessment in the validity of the finding. Even claims supported with an "extremely likely" statement should not be assumed to reflect a "high" or "very high confidence" assessment in the finding. The probabilistic output is simply one piece of evidence that may influence an author's assessment of the claim. Therefore, findings supported with likelihood statements should be paired with confidence statements assessing the validity of the finding. If there is concern that sentences combining both language scales may confuse readers, confidence statements can be used to support more general claims, while likelihood statements can be positioned as evidence supporting those claims.

Additionally, the guidance should explicitly endorse the more common formulaic interpretation of the relationship between the evidence/agreement and confidence scales and provide greater clarity on when and how evidence/agreement terms are translated into confidence terms:

Mandate that each chapter team establish its own rules for determining when and how to translate evidence/agreement statements into confidence statements at the beginning of the assessment process. These rules must be followed consistently within the chapter and should be articulated somewhere in the chapter body.

4.7.2 Training, oversight, and resources

Augment training with author-relevant examples. While time constraints may make it difficult to expand the training offered to IPCC authors on the uncertainty language framework, assigning members of the IPCC Bureau or other "uncertainty framework experts" to run through exercises with authors could significantly improve familiarity with the three scales and help authors generate questions about the framework that would typically emerge much later in the assessment process. Exercises can be enhanced by using examples that involve the specific topics and types of evidence that each author is most familiar with.

Increase IPCC Bureau and home institution support of IPCC authors. The rigorous and consistent application of the uncertainty language framework depends on there being sufficient resources, oversight, and time for IPCC authors. The IPCC Bureau can better support IPCC authors by providing more uncertainty expertise than what is typically offered by the overburdened co-Chairs. One possible solution is appointing an additional review editor to each chapter (or to multiple chapters) who is tasked with monitoring the use of uncertainty language and offering support to authors.

The most important factor for whether an author has sufficient time to produce the best assessment possible is the level of support they receive from their home institution. Institutions that offer authors meaningful support (e.g. reduced teaching responsibilities) during the final months of the assessment cycle, as well as the months immediately following report approval are significantly better equipped to uphold the high standards of the IPCC assessment process and are more likely to be able to participate in future assessments.

4.7.3 Conclusion

In Article #1, I define uncertainty as a mental state of imperfect confidence in the extent or quality of one's knowledge. This definition aligns closely with the IPCC's confidence scale where an author's confidence level is a product of their subjective assessment of the amount and quality of evidence, as well as the level of agreement among other experts. However, the guidance accompanying the IPCC uncertainty framework is careful to differentiate between qualitative confidence assessments (which, the guidance claims, do not contain probabilistic information) and quantitative probability intervals produced primarily by frequentist methods. However, from a confidence-deficit or Bayesian perspective, a subjective judgement about the truthfulness of a knowledge claim (i.e. a confidence assessment) *does* contain probabilistic information.

Conversations with IPCC Bureau members and a close examination of the IPCC uncertainty language guidance reveal that the IPCC's uneasiness about recognizing the probabilistic nature of the confidence scale reflects a desire to maintain the perception of objectivity around the IPCC assessment process. Numerical probability values produced by statistical or modeling studies are seen by policy makers as more objective and trustworthy than those produced by an author (or collectively by an author team) to reflect their subjective confidence level in the accuracy of a knowledge claim.

However, this view reflects a false dichotomy between "objective" probabilities produced by models and subjective probabilities produced by experts. As I argue in Section 4.4.2, the modeling studies underpinning the likelihood assessments in IPCC reports involve many subjective decisions, including a judgement that the model adequately reflects the real world system and that trials are sufficiently homogeneous. Further, many of the likelihood terms used in IPCC reports reflect the aggregation of multiple models, which requires authors to make a number of judgement calls around the quality and compatibility of models.

At the same time, it is important to acknowledge that these quantitative models and statistical analyses constitute some of the strongest evidence available to IPCC authors for assessing the state of knowledge on the various systems implicated in climate change. Clearly, the results of these studies should underpin the claims made in IPCC reports. However, distinguishing these findings from subjective probabilities by using separate language scales only serves to confuse authors and policy makers – and to perpetuate the myth that frequencies (necessarily) have higher epistemic value than subjective probabilities. Further, it creates confusion about whether the frequencies presented in IPCC reports reflect an author's assessment of the truthfulness of the knowledge claim (Does a 50-100% finding from a model mean that the author is 50-100% confident in that finding?) – or whether it is simply an important piece of evidence that influences the author's assessment.

Beyond the AR6 cycle, the IPCC Bureau should consider a simplification of the uncertainty language framework that is based on a more rigorous and consistent interpretation of probability. A good starting point is clarifying the status of probabilistic findings from statistical and modeling studies, like I suggest in Section 4.7.1. The link between quantitative findings expressed as probability intervals and the specific study (or studies) that produced them should be clearly described in the text. For instance, it should be clear that probability X was produced by study Y. Then, a separate confidence term should be attached to the assessment of the overarching knowledge claim (even if the confidence assessment is based entirely on the finding from the quantitative study).

More radically, I think that the IPCC should eliminate the likelihood language scale entirely. Quantitative findings from statistical and modeling analyses can still be presented as probability values or probability intervals (so long as it is clearly communicated that those numbers come from a specific study or studies) but it is unnecessary to take the further step of linking those values to a qualitative label like "very likely." Since an additional confidence term is necessary to clarify the author's assessment of the overarching claim, the likelihood term only serves to clutter and confuse the communication of quantitative evidence.

Meanwhile, the formulaic interpretation of the distinction between the evidence/agreement and confidence scales means that the two scales are essentially redundant, with the evidence/agreement scale providing slightly more fine-grained information (particularly when the conversion requires a judgement call for terms like "robust evidence, medium agreement"). The observed preference for translating evidence/agreement terms into confidence terms in the chapter executive summaries and SPM reveals that authors value concision and clarity in these more widely read sections of IPCC reports. Meanwhile, a number of interviewees expressed their preference for using the evidence/agreement scales in the chapter bodies because it allows them to "show their work." Therefore, I suggest that *only* evidence/agreement terms be used in the chapter bodies, while *only* confidence terms be used in the executive summaries and SPM.

While AR6 marks the first major assessment report that has not been preceded by an update to the IPCC uncertainty language framework in over 20 years, the implementation of the framework in the three recent SRs reveals that familiar inconsistencies and sources of confusion persist. The AR6 provides an opportunity to address many of these issues. Ultimately, I am sensitive to the argument that author familiarity with the existing framework is an important concern – which was the main justification for why the framework was not updated after AR5. However, I strongly support a reenergized conversation about more significant changes to the framework that can balance author familiarity with a more consistent and scientifically rigorous interpretation of probability – along with better support, oversight, and resources.

Chapter 5

Conclusion: Supporting uncertain policy decisions for global catastrophic risks

Each of the three articles in this dissertation contributes to a distinct interdisciplinary conversation on the theme of decision support: the conceptual literature on uncertainty, the organizational studies literature on R&D priority-setting, and the literature on climate change governance and science communication. However, this dissertation is *not* an exercise in disciplinary bridge-building. Rather, this project is more of an exercise in "normal science," adopting the specific language and conventions of each research program in order to advance three distinct but important conversations.

What unites the three papers is their central focus on decision support – the process whereby experts take actions to improve the ratio of "what we know" to "what we do not know" and then communicate them to decision makers addressing complex policy problems. Decision support has become a bottleneck for efforts to address policy problems in the 21st century – one that must be addressed from multiple angles if we have any hope of rising to the challenge posed by the confluence of severe environmental, social, economic, and technological stresses that have the potential of producing catastrophic and irreversible impacts in our lifetime.

Over the course of writing this dissertation back in 2018 and 2019, I witnessed the rise of a global political culture that is increasingly suspicious of, and hostile towards, experts and technocrats – a trend that has only been amplified by the COVID-19 pandemic. With this context in mind, I hope that my research can bolster efforts to reinforce the message that cutting-edge, well-communicated science and expertise are absolutely essential if our governance institutions are to rise to the challenge of addressing problems like climate change, NEO impacts, infectious disease outbreaks, and the misuse of dual-use technologies. My aim is to provide pragmatic and prescriptive analysis of decision support arrangements embedded within key governance organizations. Similar analysis addressing other organizations and policy problems is urgently needed. And this urgency will only become more pronounced as GCRs continue to shift from the background to the foreground of the political agenda.

I chose to focus my analysis on organizations with decision support relationships that specifically address GCRs – risks of events that could significantly harm or destroy human civilization on a global scale. For Article #2, the decision to highlight the decision support challenges associated with planetary defense governance (and GCRs in general) stems from the lack of attention paid to GCRs within the fields of organizational studies and science policy. In contrast, Article #3 explores an organization (in the IPCC) and a GCR (in climate change) that have received a relatively large amount of scholarly attention. In this case, my analysis was motivated by the urgent nature of the climate change issue and the fact that the IPCC is the primary knowledge broker in the global climate change governance regime. In both cases, GCRs provide a useful lens for analyzing key decision support tasks like uncertainty reduction and uncertainty communication because of their enormous stakes and their inextricable link to uncertainty.

5.1 Summary of findings

Article #1 examines existing conceptual frameworks of uncertainty and concludes that they are epistemologically inconsistent, redundant, and fail to live up to their stated goal of providing a comprehensive treatment of the uncertainty concept. I propose several amendments to the popular three-dimensional conceptualization of uncertainty that has been widely embraced in the environmental risk literature. The goal of these amendments is to not only address frustrating inconsistencies within conventional uncertainty conceptualizations but also to create a framework that is better aligned with the majority of scientific fields that approach uncertainty and probability from a confidence-deficit (i.e. Bayesian) perspective. This perspective is carried forward throughout the remainder of the dissertation.

Article #2 makes three contributions to the R&D priority setting literature and the fields of organizational studies and science policy more generally. First, it argues that R&D priority-setting (i.e. systematic uncertainty reduction) is a core governance activity for GCRs and other policy problems where the benefits produced by uncertainty-reducing R&D projects are, themselves, deeply uncertain. Second, it proposes a novel framework for explaining R&D priority-setting outcomes at public sector organizations. And third, it illustrates the explanatory value of the framework by applying it to a descriptive case study of R&D priority-setting processes at NASA.

Article #3 contributes to the critical literature on the IPCC's uncertainty language framework as the first analysis of the framework's application in the three most recent IPCC special reports. The paper draws on the different conceptualizations of uncertainty explored in Article #1 to investigate competing interpretations of the uncertainty language framework that have left many IPCC lead authors and readers of the reports confused and frustrated. This article proposes a series of improvements to the uncertainty communication framework that can be applied in AR6 and beyond.

5.2 What next?

Most PhD dissertations conclude by sketching out a future research program that builds off of the scholar's intellectual contributions and explores the many unresolved questions that emerge from the dissertation. There are certainly many exciting and important directions one could take the topics of decision support, decision-making under uncertainty, and GCR governance. For instance, I could dive deeper into the concept of ambiguity, investigating whether its possible to differentiate between specific types or dimensions of value disagreements and divergent knowledge frames. I could take my amended conceptual framework of uncertainty and use it to identify and resolve situations where experts and decision makers are not speaking the same "uncertainty language." I could also apply the insights from my descriptive study on R&D priority-setting to new studies but instead take a more prescriptive and normative approach. For instance, I could make recommendations to improve the quality of R&D decisions at organizations like NASA or compare the efficacy of its priority-setting processes to similar mission-oriented organizations. And with the AR6 set to be published in 2021/2022, I could extend my analysis in Article #3 and conduct a comparative analysis of the application of the IPCC's uncertainty language framework between AR6 and previous reports.

However, I am not currently planning on pursuing any of these research projects in the near future. Instead, I hope to apply my expertise on global catastrophic risk, uncertainty, uncertainty communication, and R&D priority-setting as a "decision support practitioner" within mission-oriented research organizations. I believe that my experience looking closely at how other experts confront uncertainty has equipped me with a set of useful strategies and perspectives that can help me address many of the common challenges these organizations face. Unsurprisingly, I tend to distinguish between strategies that are useful for performing uncertainty reduction tasks on the one hand and strategies for effectively communicating uncertainty to decision makers on the other. The strategies I propose here are hardly cutting-edge – but they all stem from the somewhat unique perspective on uncertainty and decision support that I advance in this dissertation.

5.2.1 Strategies for uncertainty reduction

As I explain in Article #2, mission-oriented research organizations conduct R&D that is specifically directed towards helping solve complex policy problems. Exactly what types of R&D projects organizations choose to fund depends on their mandate or scope. Research organizations tend to specialize on particular issues or approaches based on the expertise of their researchers, their beliefs about which problems impose the highest costs on society, and their estimation of the amount of benefit they can produce by addressing those problems. Of course, organizations are also influenced by the priorities of their funding sources – whether they sell their services to clients or receive funding from governments, donors, foundations, or other sources. In effect, research organizations must perform these same estimates of costs and benefits through the eyes of their funders. Sometimes these two visions align and sometimes they do not.

The research presented in this dissertation sheds light on the variety of tasks that are routinely performed by individuals with influence over decisions about which R&D projects to focus on – whether it is an individual researcher applying for research grants and mapping out their research program or a group of research managers at an organization determining how to spend next year's budget. In Article #2, I describe how problem definition is a fundamentally intersubjective process shaped by social factors. At organizations, R&D portfolio decisions are typically made collectively by a group of people (who rarely have equal input on these decisions). Whether or not these groups take time to sit down together and sketch out their understanding of the problem by constructing system maps or defining key causal relationships, all R&D decisions are based on these shared understandings of the problem.

Quite often, it is assumed that all individuals influencing R&D decisions possess an identical understanding of the key systems implicated in the problem. While this may indeed be the case for problems like NEO impacts, other risks like climate change can be framed differently depending on the perspective of the analyst. For example, in the introduction chapter, I argue that whether you view the climate system from a linear perspective or you appreciate the potential for non-linearities and tipping elements will determine whether you see the 1.5-2°C threshold as an important goal that will minimize human suffering or the edge of a cliff. How one defines the problem will influence which system elements your analysis will focus on, which solutions you consider, and your estimations of costs and benefits associated with candidate R&D projects.

Therefore, I am a strong advocate for ensuring the transparency of system understandings and problem framings within groups tasked with making R&D decisions at mission-oriented organizations. Formal system mapping exercises – particularly those that use causal mapping techniques – are a crucial tool for exposing hidden assumptions and recognizing key gaps in system understandings that need to be filled in order to make informed decisions. While some might view such exercises as a time-consuming, low-value strategy, I believe that they provide many other positive externalities such as externalizing the expertise of each participant, which can help eliminate redundant or ineffective investments in effort.

System mapping exercises are equally valuable when making estimates of uncertainty reducibility. In Article #1, I explain that estimating how much effort is needed to reduce a particular "decision-impeding" uncertainty is an exercise in building two types of models. First, a group of R&D portfolio managers must build a model of the evidence that could possibly increase their confidence. Take the example of trying to decrease uncertainty around whether there exist solutions to addressing the climate crisis that are both feasible (given the current state of our political, technological, economic, and social systems) and effective (meaning that they actually make a noticeable dent in addressing the problem). Today, many experts on climate change policy lack confidence that there exist interventions that check both boxes. In order to increase our confidence about whether these "high-leverage intervention points" exist, we necessarily begin by building an evidence model. Pieces of relevant evidence might include an analysis of the inadequacies of existing climate change interventions or an analysis of other system transformations that have similar characteristics to climate change like the abolishment of slavery. I strongly suggest that these evidence models, as well as subsequent models that estimate how difficult it will be to collect each piece of evidence (evidence-effort models), should also be formally mapped in group exercises. Again, transparency brings assumptions to the surface and fosters a collaborative and reflexive organizational culture.

Lastly, I am also a proponent of establishing standard processes that help guide R&D portfolio managers through the main stages of the R&D priority-setting process outlined in Article #2 (defining the problem, objectives, decision points, and uncertainties; estimating benefits and costs; conducting CBA; and selecting projects). While I recognize that these stages are iterative and overlapping, I believe that each task should be addressed somewhat individually, which is not to say that each task needs its own two-day system mapping workshop. Rather, simply using the decision analysis cycle as a process framework to track the evolution of annual R&D decisions can help to emphasize the intersubjective nature of these decisions and the implicit modeling tasks performed at each stage, which will ultimately result in more intentional and transparent decisions.

5.2.2 Strategies for communicating uncertainty

Mission-oriented organizations need to produce research that is seen as scientifically rigorous by other experts while simultaneously seen as legitimate and useful by decision makers. When the IPCC began producing assessment reports 30 years ago, the main threat to the perceived legitimacy of early assessment reports came from skeptics that used expressions of uncertainty in the reports to undermine the scientific consensus on key findings. Even today, the increasingly prominent presentation of uncertainty language in IPCC reports has been described as "self-defeating" (Meah 2019), while studies show that uncertainty language confuses certain audiences (Hollin and Pearce 2015; Rabinovich and Morton 2012).

However, other commentators argue that the IPCC should focus less on self-protection from skeptics and a move towards a "full exploration of uncertainty" (Oppenheimer et al. 2007, pp. 1506), reflecting a desire for a greater emphasis on transparently communicating the expert judgements that underpin knowledge claims. However, the honest and transparent communication of uncertainty may be at odds with the preferences of policy makers who tend to gravitate towards "high certainty" findings (Meah 2019).

The balancing act between protecting science and expert knowledge from skeptics while honestly communicating knowledge gaps is an inherent trade-off in any uncertainty communication framework. I believe that the way that uncertainty is communicated for any knowledge claim should reflect the specific purpose of the decision support task. In other words, uncertainty communication must be "fit for purpose." In one case, fit for purpose could mean unpacking the underlying evidence base and the nature of uncertainty in great detail. In another case, it could mean making a knowledge claim with a fairly straightforward declaration of confidence and then pointing to further material that readers can explore if they are interested in investigating that claim further. One lesson that emerged from both the NASA and IPCC cases is that some level of formal or informal peer review – having multiple experts "weigh in" on an uncertainty assessment – is a crucial strategy for inserting an element of quality control into the uncertainty characterization and communication process. Intersubjective assessments made by more than one expert will always be granted more legitimacy than subjective individual assessments.

The other strategy that emerged from my critical analysis of conceptual frameworks of uncertainty in Article #1 and the interviews I conducted with IPCC lead authors in Article #3 was the importance of clear and consistent definitions of key concepts. While I argue in Article #1 that ideally, concepts relating to uncertainty should be logically consistent and non-redundant, in an organizational context like the IPCC, it is far more important that concepts be used consistently, no matter how they are defined.

For example, in Article #3, I propose two pathways the IPCC could take to significantly improve its uncertainty language framework. First, I provide recommendations that address the application of the IPCC uncertainty language framework on its own terms. My primary aim here is to provide pragmatic recommendations to clarify the existing framework with the hope that they could be applied to AR6, as well as subsequent special reports that may be initiated before a new framework can be introduced.

Second, I propose that the IPCC eventually transitions to a simpler framework by eliminating the likelihood scale and embracing a system that reflects that widespread Bayesian interpretation of probability. All mission-oriented research organizations disseminate uncertain knowledge claims to decision makers and therefore, it is important that they use consistent uncertainty language – whether it is expressions of confidence, likelihood, probability, or ignorance.

To summarize, the four strategies I highlight for improving uncertainty reduction and communication by research managers at mission-oriented research organizations are:

1. Use collaborative system mapping tools and exercises to promote the transparent discussion of problem framings and system understandings, as well as the systematic exploration of the costs and benefits of R&D projects.
2. Use the decision analysis cycle as a process framework to systematize R&D priority-setting.
3. Ensure that the language used to communicate uncertain knowledge claims is "fit for purpose," comprehensible, useful, and consistent.
4. Incorporate formal or informal peer review into the process of qualifying uncertain knowledge claims wherever possible.

5.2.3 Reflections from the beginning and end of a pandemic

I am concluding this dissertation with some reflections that I wrote back in March 2020 as the novel coronavirus was sweeping through China and much of Europe and case counts were starting to climb in Canada. Most of my reflections on the state of global risk management, science communication, and our prospects for meeting the challenges of the 21st century still ring true today. On the one hand, the COVID-19 pandemic has illustrated the absolute necessity of expertise and decision support for navigating humanity's converging crises. In fact, early in the pandemic, many commentators expressed hope that the leadership role played by public health experts in the pandemic response might help rebuild the frayed relationship between large segments of society and institutions of expertise. Yet, it seems that levels of distrust, hyper-partisanship, and epistemic fragmentation have never been greater. This dissertation proposes a series of improvements to existing decision support systems—a modest but important task. While improving the quality of decision support systems is certainly one part of the solution to improving our capacity to address complex policy problems, there are certainly much larger social, cultural, and psychological obstacles we need to overcome to produce decision support systems capable of navigating us through this century.

March 13, 2020

As I sit down to write the final paragraphs of this dissertation, I think that it is worthwhile to reflect on some of the ways that the world has and has not changed since I began this project. Back in the spring of 2018, the sense of urgency around the impending climate emergency was already at a fever pitch (at least within my professional and personal networks). Facing a 12-year window to fundamentally transform our global energy system by 2030 – along with the economic, political, social, and ideological systems that underpin it – a feeling of hopelessness was beginning to set in. Now, with less than 10 years left on the clock, the situation does not look much different.

However, we have seen a few encouraging developments. The student strikes, led by the youth climate activist Greta Thunberg, have galvanized the climate justice movement and more people around the world are aware (and fearful) of climate change today than ever before. Meanwhile, the appetite of banks and large investors to continue financing oil and gas projects appears to be slowly diminishing. However, few countries are on track to come anywhere close to meeting their Paris commitments, which were lackluster to begin with.

But today, the world's attention is firmly placed on the unfolding COVID-19 pandemic. As I write, there have been over 81,000 confirmed cases and over 3,200 deaths. China and Italy have completely shut down their borders and have essentially suspended all economic activity. Around the world, sporting events, conferences, and other mass gatherings are being widely cancelled and individuals are being encouraged to practice social distancing measures to slow the rate of the spread. Last week, the Canadian Prime Minister's wife tested positive for the virus and the first case was declared in the city of Kitchener where my wife Amy and I live. While public health officials are hopeful that the outbreak will subside over the summer, the only real way to control the spread is with the development and deployment of an effective vaccine – which is not expected to arrive until next year at the earliest.

Epidemiologists are predicting anywhere from 30-70% of the global population will be infected when all is said and done with a case fatality rate somewhere between 0.2% and 3%. Even the worst case scenario would fall below the threshold of a truly "catastrophic" risk event but would serve as a sobering wake up call to the inevitability of such outbreaks and the vulnerability of our social, economic, and political systems to effectively respond to them. If this pandemic is a dress rehearsal for responding to a far more catastrophic outbreak in the future, I believe it is safe to say that we have failed the test.

While it hardly constitutes a silver lining, the year 2020 will likely see the first dip in global carbon emissions since the 2008-2009 economic crisis. Unfortunately, these decreases will almost certainly be short-lived once the economic engines get turned back on and governments try to walk the global economy back from the brink.

It remains to be seen which lessons we will take away from the pandemic once it runs its course. Optimistically, I hope that we will recognize the truly global nature (in both cause and effect) of problems like infectious disease outbreaks and climate change, and that we will recognize the necessity of pulling together as a species to solve these kinds of problems together. I also hope that we will see the necessity of seriously and systematically thinking about our uncertain future – and taking precautionary measures to restructure and fortify our systems against unexpected shocks. And most of all, I hope that we will see the mending of our crumbling relationship with institutions of expertise, recognizing that our very survival may depend on placing our trust in institutions and individuals singularly focused on confronting the uncertainty that paralyzes good decision making. Because the world could use good decisions more than ever.

Bibliography

Adalja, Amesh A, Matthew Watson, Eric S Toner, Anita Cicero, and Thomas V Inglesby. 2018. "The Characteristics of Pandemic Pathogens." Baltimore, MD. http://www.centerforhealthsecurity.org/our-work/pubs_archive/pubs-pdfs/2018/180510-pandemic-pathogens-report.pdf.

Adler, Carolina E., and Gertrude Hirsch Hadorn. 2014. "The IPCC and Treatment of Uncertainties: Topics and Sources of Dissensus." *Wiley Interdisciplinary Reviews: Climate Change* 5: 663–76. https://doi.org/10.1002/wcc.297.

Alvarez, L. W., W. Alvarez, F. Asaro, and H. V. Michel. 1980. "Extraterrestrial Cause for the Cretaceous-Tertiary Extinction." *Science* 208: 1095–1108.

Aronowitz, S. 1988. *Science as Power: Discourse and Ideology in Modern Society*. Minneapolis: University of Minnesota Press.

Arratia, M. N.M., I. F. López, S. E. Schaeffer, and L. Cruz-Reyes. 2016. "Static R&D Project Portfolio Selection in Public Organizations." *Decision Support Systems* 84: 53–63. https://doi.org/10.1016/j.dss.2016.01.006.

Ascough II, J. C., H. R. Maier, J. K. Ravalico, and M. W. Strudley. 2008. "Future Research Challenges for Incorporation of Uncertainty in Environmental and Ecological Decision-Making." *Ecological Modelling* 219 (3–4): 383–99. https://doi.org/10.1016/j.ecolmodel.2008.07.015.

Asher, D. J., M. Bailey, V. Emel'yanenko, and W. Napier. 2005. "Earth in the Cosmic Shooting Gallery." *The Observatory* 125: 319–22.

Asselt, Marjolein B. A. van, and Jan Rotmans. 2002. "Uncertainty in Integrated Assessment Modelling: From Positivism to Pluralism." *Climatic Change* 54: 75–105.

Asselt, Marjolein B.A. van. 2000. *Perspectives on Uncertainty and Risk: The PRIMA Approach to Decision Support*. Dordecht, The Netherlands: Kluwer Academic Publishers.

Atkinson, A. 1999. *Impact Earth: Asteroids, Comets and Meteors—the Growing Threat*. London: Virgin.

Aven, Terje. 2019. "Climate Change Risk–What Is It and How Should It Be Expressed?" *Journal of Risk Research*, 1–18. https://doi.org/10.1080/13669877.2019.1687578.

Aven, Terje, and Ortwin Renn. 2015. "An Evaluation of the Treatment of Risk and Uncertainties in the IPCC Reports on Climate Change." *Risk Analysis* 35 (4): 701–12. https://doi.org/10.1111/risa.12298.

Avin, Shahar. 2014. "Breaking the Grant Cycle: On the Rational Allocation of Public Resources to

Scientific Research Projects.” University of Cambridge.

Avin, Shahar, Bonnie C. Wintle, Julius Weitzdörfer, Seán S. Ó hÉigeartaigh, William J. Sutherland, and Martin J. Rees. 2018. “Classifying Global Catastrophic Risks.” *Futures* 102 (February): 20–26. https://doi.org/10.1016/j.futures.2018.02.001.

Baecher, GB, and JC Christian. 2000. “Natural Variation, Limited Knowledge, and the Nature of Uncertainty in Risk Analysis.” *Risk and Uncertainty in Water Resources*, no. 2: 1–16. http://pm.umd.edu/files/public/water_library/2000/Natural Variation, Limited Knowledge, and the Nature of Uncertainty in Risk Analysis_2000_Baecher.pdf.

Barnett, M., and R. Duvall. 2005. “Power in International Politics.” *International Organization* 59 (1): 39–75.

Barnosky, Anthony D., Nicholas Matzke, Susumu Tomiya, Guinevere O. U. Wogan, Brian Swartz, Tiago B. Quental, Charles Marshall, et al. 2011. “Has the Earth’s Sixth Mass Extinction Already Arrived?” *Nature* 471 (7336): 51–57.

Barrett, Anthony Michael. 2017. “Value of Global Catastrophic Risk (GCR) Information: Cost-Effectiveness-Based Approach for GCR Reduction.” *Decision Analysis* 14 (3): 187–203. https://doi.org/10.1287/deca.2017.0350.

Barrett, Anthony Michael, and Seth D. Baum. 2017. “Risk Analysis and Risk Management for the Artificial Superintelligence Research and Development Process.” In *The Technological Singularity: Managing the Journey*, edited by V. Callaghan and J. Miller. Berlin: Springer.

Bates, Matthew E., Jeffrey M. Keisler, Niels P. Zussblatt, Kenton J. Plourde, Ben A. Wender, and Igor Linkov. 2016. “Balancing Research and Funding Using Value of Information and Portfolio Tools for Nanomaterial Risk Classification.” *Nature Nanotechnology* 11 (2): 198–203. https://doi.org/10.1038/nnano.2015.249.

Baum, Seth D. 2010. “Is Humanity Doomed? Insights from Astrobiology.” *Sustainability* 2 (2): 591–603. https://doi.org/10.3390/su2020591.

———. 2019. “Accounting For Violent Conflict Risk In Planetary Defense Decisions.” In *Proceedings of the 6th IAA Planetary Defense Conference*.

Baum, Seth D., and Itsuki C. Handoh. 2014. “Integrating the Planetary Boundaries and Global Catastrophic Risk Paradigms.” *Ecological Economics* 107: 13–21.

Baustert, Paul, Benoit Othoniel, Benedetto Rugani, and Ulrich Leopold. 2018. “Uncertainty Analysis in Integrated Environmental Models for Ecosystem Service Assessments: Frameworks, Challenges and Gaps.” *Ecosystem Services* 33: 110–23. https://doi.org/10.1016/j.ecoser.2018.08.007.

Beck, Silke, and Martin Mahony. 2018. “The IPCC and the New Map of Science and Politics.” *Wiley Interdisciplinary Reviews: Climate Change* 9: e547. https://doi.org/10.1002/wcc.547.

Bedford, T., and R. M. Cooke. 2001. *Probabilistic Risk Analysis: Foundations and Methods*. Cambridge, UK: Cambridge University Press.

Beer, Tom. 2006. “Ecological Risk Assessment and Quantitative Consequence Analysis.” *Human and Ecological Risk Assessment* 12 (1): 51–65. https://doi.org/10.1080/10807030500428587.

Beinhocker, Eric D. 2013. “Reflexivity, Complexity, and the Nature of Social Science.” *Journal of Economic Methodology*. Taylor & Francis. https://doi.org/10.1080/1350178X.2013.859403.

Bernstein, Steven. 2001. *The Compromise of Liberal Environmentalism*. New York: Columbia University Press.

Bethe, H. A. 1976. “Ultimate Catastrophe?” *Bulletin of the Atomic Scientists* 32 (6): 36–37.

Beven, Keith. 2016. “Facets of Uncertainty: Epistemic Uncertainty, Non-Stationarity, Likelihood, Hypothesis Testing, and Communication.” *Hydrological Sciences Journal* 61 (9): 1652–65. https://doi.org/10.1080/02626667.2015.1031761.

Bhattacharjya, Debarun, Jo Eidsvik, and Tapan Mukerji. 2013. “The Value of Information in Portfolio Problems with Dependent Projects.” *Decision Analysis* 10 (4): 341–51.

Biermann, Frank. 2012. “Planetary Boundaries and Earth System Governance: Exploring the Links.” *Ecological Economics* 81: 4–9. https://doi.org/10.1016/j.ecolecon.2012.02.016.

Borges de Amorim, Pablo, and Pedro B. Chaffe. 2019. “Towards a Comprehensive Characterization of Evidence in Synthesis Assessments: The Climate Change Impacts on the Brazilian Water Resources.” *Climatic Change* 155: 37–57. https://doi.org/10.1007/s10584-019-02430-9.

Bostrom, N. 2009. “Pascal’s Mugging.” *Analysis* 69 (3): 443–45. https://doi.org/10.1093/analys/anp062.

Bostrom, Nick. 2002. “Existential Risks: Analyzing Human Extinction Scenarios and Related Hazards.” *Journal of Evolution and Technology* 9: 1–37. https://doi.org/10.1080/01140671.2003.9514271.

———. 2014. *Superintelligence: Paths, Dangers, Strategies*. Oxford: Oxford University Press.

Bostrom, Nick, and Milan M. Ćirković, eds. 2008. *Global Catastrophic Risks*. Oxford, UK: Oxford University Press.

Bradley, Richard, and Mareile Drechsler. 2013. “Types of Uncertainty.” *Erkenntnis* online: 1–29. https://doi.org/10.1007/s10670-013-9518-4.

Brattström, Erik, and Tomas Hellström. 2019. “Street-Level Priority-Setting: The Role of Discretion in Implementation of Research, Development, and Innovation Priorities.” *Energy Policy* 127: 240–47. https://doi.org/10.1016/j.enpol.2018.12.006.

Briggs, David J., Clive E. Sabel, and Kayoung Lee. 2009. "Uncertainty in Epidemiology and Health Risk and Impact Assessment." *Environmental Geochemistry and Health* 31: 189–203. https://doi.org/10.1007/s10653-008-9214-5.

Brouwer, R., and C. De Blois. 2008. "Integrated Modelling of Risk and Uncertainty Underlying the Cost and Effectiveness of Water Quality Measures." *Environmental Modelling and Software* 23: 922–37.

Brown, James D. 2004. "Knowledge, Uncertainty and Physical Geography: Towards the Development of Methodologies for Questioning Belief." *Transactions of the Institute of British Geographers* 29 (3): 367–81. https://doi.org/10.1111/j.0020-2754.2004.00342.x.

Brugnach, M., Art Dewulf, Claudia Pahl-wostl, and Tharsi Taillieu. 2008. "Toward a Relational Concept of Uncertainty: About Knowing Too Little, Knowing Too Differently, and Accepting Not to Know." *Ecology and Society* 13 (2): 2010–2010.

Brugnach, M., and H. Ingram. 2012. "Ambiguity: The Challenge of Knowing and Deciding Together." *Environmental Science and Policy* 15: 60–71. https://doi.org/10.1016/j.envsci.2011.10.005.

Butzer, K. W., and G. H. Endfield. 2012. "Critical Perspectives on Historical Collapse." *Proceedings of the National Academy of Sciences* 109 (10): 3628–31. https://doi.org/10.1073/pnas.1114772109.

Campbell, K, J Gulledge, JR McNeill, J Podesta, P Ogden, L Fuerth, J Woolsley, et al. 2007. *The Age of Consequences: The Foreign Policy and National Security Implications of Global Climate Change*. Washington, D.C.: Centre for Strategic and International Studies & Centre for New American Security.

Carnap, Rudolf. 1950. *Logical Foundations of Probability*. London: Routledge & Kegan Paul.

Carter, B. 1983. "The Anthropic Principle and Its Implications for Biological Evolution." *Philosophical Transactions of The Royal Society A Mathematical Physical and Engineering Sciences* 310 (1512): 347–63.

Casti, J. L. 1996. "The Outer Limits: In Search of the 'Unknowable' in Science." In *Boundaries and Barriers: On the Limits to Scientific Knowledge*, edited by J.L. Casti and A. Karlqvist, 12–35. Reading, Massachusetts: Addison-Wesley.

Chapman, Clark R. 1999. "History of The Asteroid/Comet Impact Hazard." Boulder, Colorado. https://www.boulder.swri.edu/clark/ncarhist.html.

Chapman, Clark R. 2004. "The Hazard of Near-Earth Asteroid Impacts on Earth." *Earth and Planetary Science Letters* 222: 1–15. https://doi.org/10.1016/j.epsl.2004.03.004.

Chen, Chi-Feng, Hwong-wen Ma, and Kenneth H. Reckhow. 2007. "Assessment of Water Quality Management with a Systematic Qualitative Uncertainty Analysis." *Science of the Total Environment*

374: 13–25. https://doi.org/10.1016/j.scitotenv.2006.12.027.

Chesley, Steven R, Paul W Chodas, Andrea Milani, and Giovanni B Valsecchi. 2002. "Quantifying the Risk Posed by Potential Earth Impacts" 432: 423–32. https://doi.org/10.1006/icar.2002.6910.

Ciarli, Tommaso, and Ismael Ràfols. 2019. "The Relation between Research Priorities and Societal Demands: The Case of Rice." *Research Policy* 48: 949–67. https://doi.org/10.1016/j.respol.2018.10.027.

Clark, Stuart. 2020. "European Space Agency Awards Hera Asteroid Mission to German Firm." *The Guardian*, September 17, 2020. https://www.theguardian.com/science/2020/sep/17/european-space-agency-awards-asteroid-mission-german-firm.

Claxton, Karl. 2008. "Discussion on the Paper by Welton, Ades, Caldwell and Peters." *Journal of the Royal Statistical Society. Series A: Statistics in Society* 171 (4): 834–35.

Climate Action Tracker. 2015. "Climate Pledges Will Bring 2.7°C of Warming, Potential for More Action." *Climate Action Tracker*, December 8, 2015. https://climateactiontracker.org/publications/climate-pledges-will-bring-27c-of-warming-potential-for-more-action/.

Climenhaga, Nevin. 2019. "The Concept of Probability Is Not as Simple as You Think." *Aeon*, February 26, 2019. https://aeon.co/ideas/the-concept-of-probability-is-not-as-simple-as-you-think.

Committee to Review the Intergovernmental Panel on Climate Change. 2010. "Climate Change Assessments: Review of the Processes and Procedures of the IPCC." https://archive.ipcc.ch/pdf/IAC_report/IAC Report.pdf.

Corballis, Michael C. 2011. *The Recursive Mind: The Origins of Human Language, Thought, and Civilization*. Princeton, New Jersey: Princeton University Press.

Cotton-Barratt, Owen Sebastian Farquhar, John Halstead, Stefan Schubert, and Andrew Snyder-Beattie. 2016. "Global Catastrophic Risks 2016." https://doi.org/10.1111/j.1539-6924.2008.01162.x.

Crutchfield, J. P., J. D. Farmer, N. H. Packard, and R. S. Shaw. 1986. "Chaos." *Scientific American* 255: 46–57.

Cruz-Castro, Laura, Koen Jonkers, and Luis Sanz-Menéndez. 2015. "The Internationalization of Research Institutes." In *Towards European Science Dynamics and Policy of an Evolving European Research Space*, edited by Linda Wedlin and Maria Nedeva, 175–98. Cheltenham, UK: Edward Elgar.

Cruz-Castro, Laura, and Luis Sanz-Menéndez. 2018. "Autonomy and Authority in Public Research Organisations: Structure and Funding Factors." *Minerva* 56: 135–60. https://doi.org/10.1007/s11024-018-9349-1.

Cunha, Danilo, Rafael Xavier, and Leandro Nunes de Castro. 2018. “Bacterial Colonies as Complex Adaptive Systems.” *Natural Computing* 17: 781–98. https://doi.org/10.1007/s11047-018-9689-7.

Currie, Adrian, and Seán Ó hÉigeartaigh. 2018. “Working Together to Face Humanity’s Greatest Threats: Introduction to the Future of Research on Catastrophic and Existential Risk.” *Futures* 102: 1–5. https://doi.org/10.1016/j.futures.2018.07.003.

Curry, Judith. 2011. “Reasoning about Climate Uncertainty.” *Climatic Change* 108: 723–32. https://doi.org/10.1007/s10584-011-0180-z.

D’Este, Pablo, Irene Ramos-Vielba, Richard Woolley, and Nabil Amara. 2018. “How Do Researchers Generate Scientific and Societal Impacts? Toward an Analytical and Operational Framework.” *Science and Public Policy* 45 (6): 752–63. https://doi.org/10.1093/SCIPOL/SCY023.

Dafoe, Allan. 2018a. “AI Governance: A Research Agenda.” Oxford, UK. https://www.fhi.ox.ac.uk/wp-content/uploads/GovAIAgenda.pdf.

———. 2018b. “Governing the AI Revolution : A Research Agenda for Political Science.”

Dar, A., A. De Rujula, and U. Heinz. 1999. “Will Relativistic Heavy-Ion Colliders Destroy Our Planet?” *Physics Letters B* 470: 142.

Dequech, David. 2011. “Uncertainty: A Typology and Refinements of Existing Concepts.” *Journal of Economic Issues* 0 (3): 621–40. https://doi.org/10.2753/JEI0021-3624450306.

Dewulf, Art. 2013. “Contrasting Frames in Policy Debates on Climate Change Adaptation.” *WIREs Climate Change* 4: 321–30. https://doi.org/10.1002/wcc.227.

Dewulf, Art, Marc Craps, Rene Bouwen, Tharsi Taillieu, and C. Pahl-Wostl. 2005. “Integrated Management of Natural Resources: Dealing with Ambiguous Issues, Multiple Actors and Diverging Frames.” *Water Science and Technology* 52 (6): 115–24. https://doi.org/10.1016/j.diabres.2005.05.017.

Drago, Carlo, and Matteo Ruggeri. 2019. “Setting Research Priorities in the Field of Emergency Management: Which Piece of Information Are You Willing to Pay More?” *Quality & Quantity* 53: 2103–15. https://doi.org/10.1007/s11135-019-00858-w.

Dreier, Casey. 2019. “How NASA’s Planetary Defense Budget Grew By More Than 4000% in 10 Years.” *The Planetary Society*, September 26, 2019. http://www.planetary.org/blogs/casey-dreier/2019/nasas-planetary-defense-budget-growth.html.

Edenhofer, Ottmar, and Martin Kowarsch. 2012. “A Pragmatist Concept of Science Policy Advice.” Berlin.

Edwards, Ward, Ralph F. Miles Jr., and Detlof von Winterfeldt. 2007. “Introduction.” In *Advances in*

Decision Analysis: From Foundations to Applications, edited by Ward Edwards, Ralph F. Miles Jr., and Detlof von Winterfeldt, 1–12. Cambridge, UK: Cambridge University Press.

Elith, Jane, Mark A Burgman, and Helen M Regan. 2002. "Mapping Epistemic Uncertainties and Vague Concepts in Predictions of Species Distribution." *Ecological Modelling* 157: 313–29. www.elsevier.com/locate/ecolmodel.

Evershed, Nick, Andy Ball, and Naaman Zhou. 2020. "How Big Are the Fires Burning in Australia? Interactive Map." *The Guardian*, January 24, 2020. https://www.theguardian.com/australia-news/datablog/ng-interactive/2019/dec/07/how-big-are-the-fires-burning-on-the-east-coast-of-australia-interactive-map.

Faber, Malte, Reiner Manstetten, and John L.R. Proops. 1992. "Humankind and the Environment: An Anatomy of Surprise and Ignorance." *Environmental Values* 1 (3): 217–41. https://doi.org/10.3197/096327192776680089.

Fernholz, Tim. 2019. "NASA Won't Launch a Mission to Hunt Deadly Asteroids." *Quartz*, July 5, 2019. https://qz.com/1659566/nasa-nixes-hunt-for-deadly-asteroids/.

Finetti, Bruno de. 1931. "Sul Significato Soggettivo Della Probabilit`a." *Fundamenta Mathematicae* 17: 298–329.

———. 1937. "La Prévision: Ses Lois Logiques, Ses Sources Subjectives." *Annales de l'Institut Henri Poincaré* 7: 1–68.

Finkel, Adam M. 1990. "Confronting Uncertainty in Risk Management." Washington, D.C.

Finnemore, M. 1996. "Norms, Culture, and World Politics: Insights from Sociology's Institutionalism." *International Organization* 50 (2): 325–47.

Frank, Morgan R., David Autor, James E. Bessen, Erik Brynjolfsson, Manuel Cebrian, David J. Deming, Maryann Feldman, et al. 2019. "Toward Understanding the Impact of Artificial Intelligence on Labor." *Proceedings of the National Academy of Sciences* 116 (14): 6531–39.

Fuller, Ted, and Lorraine Warren. 2006. "Entrepreneurship as Foresight: A Complex Social Network Perspective on Organisational Foresight." *Futures* 38: 956–71. https://doi.org/10.1016/j.futures.2005.12.016.

Garshnek, Victoria, David Morrison, and Frederick M Burkle. 2000. "The Mitigation, Management, and Survivability of Asteroid/Comet Impact with Earth." *Space Policy* 16: 213–22. https://ac-els-cdn-com.proxy.lib.uwaterloo.ca/S0265964600000254/1-s2.0-S0265964600000254-main.pdf?_tid=abde3ecb-8d3d-439a-a429-b59afa78b8e9&acdnat=1548259088_dce264aeac22ec6fe41c934cb9b81b37.

Gasparrini, Antonio, Yuming Guo, Francesco Sera, Ana Maria Vicedo-Cabrera, Veronika Huber, Shilu Tong, Micheline de Sousa Zanotti Stagliorio Coelho, et al. 2017. "Projections of Temperature-Related Excess Mortality under Climate Change Scenarios." *The Lancet Planetary Health* 1 (9): e360–67. https://doi.org/10.1016/S2542-5196(17)30156-0.

Global Catastrophic Risk Institute (GCRI). 2018. "GCR Concept Project." 2018. http://gcrinstitute.org/concept/.

Global Challenges Foundation. 2018. "Global Catastrophic Risks 2018." Stockholm.

Gog, Julia R., Lorenzo Pellis, James L.N. Wood, Angela R. McLean, Nimalan Arinaminpathy, and James O. Lloyd-Smith. 2015. "Seven Challenges in Modeling Pathogen Dynamics Within-Host and across Scales." *Epidemics* 10: 45–48. https://doi.org/10.1016/j.epidem.2014.09.009.

Goldman, A. 1998. "Reliabilism." In *The Routledge Encyclopedia of Philosophy*. Taylor & Francis. https://plato.stanford.edu/archives/win2016/entries/reliabilism/.

Good, I. J. 1971. "46656 Varieties of Bayesians." *American Statistician* 25 (5): 62–63.

Grace, Katja, John Salvatier, Allan Dafoe, Baobao Zhang, and Owain Evans. 2018. "Viewpoint: When Will AI Exceed Human Performance? Evidence from AI Experts." *Journal of Artificial Intelligence Research* 62: 729–54. https://doi.org/10.1613/jair.1.11222.

Haas, Peter M. 1992. "Introduction: Epistemic Communities and International Policy Coordination." In *Knowledge, Power, and International Policy Coordination*, edited by Peter M. Haas, 1–36. Columbia, S.C.: University of South Carolina Press.

Habermas, J. 1971. *Toward a Rational Society*. Boston: Beacon Press.

Hartmann, Stephan, and Jan Sprenger. 2011. "Bayesian Epistemology." In *The Routledge Companion to Epistemology*, edited by Sven Bernecker and Duncan Pritchard, 609–20. New York: Routledge.

Hayes, K., H. Regan, M. Burgman, R. Devlin, A. Kapuscinski, and J. Dambacher. 2006. "Methods to Address Uncertainty in Ecological Risk Assessment of Genetically Modified Organisms." In *Ninth International Symposium on the Biosafety of Genetically Modified Organisms*. Jeju Island, South Korea.

Hazen, Gordon B., and Jayavel Sounderpandian. 1999. "Lottery Acquisition versus Information Acquisition: Prices and Preference Reversals." *Journal of Risk and Uncertainty* 18 (2): 125–36. https://doi.org/10.1023/A:1007834413032.

Heberling, Tamra, Galen Gisler, Catherine Plesko, and Robert Weaver. 2017. "Calculating the Momentum Enhancement Factor for Asteroid Deflection Studies." *Procedia Engineering* 204: 124–29. https://doi.org/10.1016/j.proeng.2017.09.764.

Helgeson, Casey, Richard Bradley, and Brian Hill. 2018. "Combining Probability with Qualitative Degree-of-Certainty Metrics in Assessment." *Climatic Change* 149: 517–25. https://doi.org/10.1007/s10584-018-2247-6.

Hellström, Tomas, Merle Jacob, and Karolin Sjöö. 2017. "From Thematic to Organizational Prioritization: The Challenges of Implementing RDI Priorities." *Science and Public Policy* 44 (5): 599–608. https://doi.org/10.1093/scipol/scw087.

Helton, J. C. 1994. "Treatment of Uncertainty in Performance Assessments for Complex Systems." *Risk Analysis* 14 (4): 483–511.

Hempsell, C. M. 2004a. "The Investigation of Natural Catastrophes." *Journal of the British Interplanetary Society* 75: 2–13.

———. 2004b. "The Investigation of Natural Global Catastrophes." *JBIS - Journal British Interplanetary Society* 57: 2–13.

Hettne, Björn. 2005. "Beyond the 'new' Regionalism." *New Political Economy* 10 (4): 543–71.

Higham, Tom, Katerina Douka, Rachel Wood, Christopher Bronk Ramsey, Fiona Brock, Laura Basell, Marta Camps, et al. 2014. "The Timing and Spatiotemporal Patterning of Neanderthal Disappearance." *Nature* 512 (7514): 306–9.

Hildebrand, A. R., G. T. Penfield, D. A. Kring, M. Pilkington, A. Camargo, S. B. Jacobsen, and W. V. Boynton. 1991. "Chicxulub Crater: A Possible Cretaceous/Tertiary Boundary Impact Crater on the Yucatán Peninsula, Mexico." *Geology* 19 (9): 867–71.

Hildenbrandt, H., C. Carere, and C. K. Hemelrijk. 2010. "Self-Organized Aerial Displays of Thousands of Starlings: A Model." *Behavioral Ecology* 21: 1349–59. https://doi.org/10.1093/beheco/arq149.

Hill, Brian. 2019. "Confidence in Belief , Weight of Evidence and Uncertainty Reporting." *Proceedings of Machine Learning Research* 103: 235–45.

Hollin, G. J.S., and W. Pearce. 2015. "Tension between Scientific Certainty and Meaning Complicates Communication of IPCC Reports." *Nature Climate Change* 5: 753–56. https://doi.org/10.1038/nclimate2672.

Homer-Dixon, Thomas. 2000. *The Ingenuity Gap*. Toronto: Vintage Canada.

Homer-Dixon, Thomas, Jonathan Leader Maynard, Matto Mildenberger, Manjana Milkoreit, Steven J. Mock, Stephen Quilley, Tobias Schröder, and Paul Thagard. 2013. "A Complex Systems Approach to the Study of Ideology: Cognitive-Affective Structures and the Dynamics of Belief Systems." *Journal of Social and Political Psychology* 1 (1): 337–63. https://doi.org/10.5964/jspp.v1i1.36.

Howard, Ronald A. 2007. *The Foundations of Decision Analysis*. Edited by Ward Edwards, Ralph F.

Miles Jr., and Detlof von Winterfeldt. *Advances in Decision Analysis: From Foundations to Applications*. Cambridge, UK: Cambridge University Press. https://doi.org/10.1109/TSSC.1968.300115.

IPCC. 2005. “Guidance Notes for Lead Authors of the IPCC Fourth Assessment Report on Addressing Uncertainties.” Geneva, Switzerland. https://doi.org/10.1016/B978-0-12-375067-9.00128-5.

———. 2014. *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Edited by Core Writing Team, R.K. Pachauri, and L.A. Meyer. Geneva, Switzerland: IPCC.

———. 2018. “Summary for Policymakers.” In *Global Warming of 1.5°C. An IPCC Special Report on the Impacts of Global Warming of 1.5°C above Pre-Industrial Levels and Related Global Greenhouse Gas Emission Pathways, in the Context of Strengthening the Global Response to the Threat of Climate Change,* edited by V. Masson-Delmotte, P. Zhai, H. O. Pörtner, D. Roberts, J. Skea, P. R. Shukla, A. Pirani, et al. Geneva, Switzerland: World Meteorological Organization.

———. 2019a. “IPCC Authors (Beta).” 2019. https://archive.ipcc.ch/report/authors/.

———. 2019b. “Summary for Policymakers.” In *IPCC Special Report on the Ocean and Cryosphere in a Changing Climate*, edited by H.O. Pörtner, D.C. Roberts, V. Masson-Delmotte, P. Zhai, M. Tignor, E. Poloczanska, K. Mintenbeck, et al. In press.

———. 2019c. “Summary for Policymakers (Approved Draft).” In *IPCC Special Report Climate Change, Desertification, Land Degradation, Sustainable Land Management, Food Security, and Greenhouse Gas Fluxes in Terrestrial Ecosystems*. In press. https://www.ipcc.ch/site/assets/uploads/2019/08/Edited-SPM_Approved_Microsite_FINAL.pdf.

Irvine, John, and Ben R. Martin. 1984. *Foresight in Science. Picking the Winners*. London: Pinter.

Jablonowski, M. 2007. *Risk Dilemmas: Forced Choices and Survival*. New York: Palgrave Macmillan.

Jaffe, R. L., W. Busza, J. Sandweiss, and F. Wilczek. 1999. “Review of Speculative ‘Disaster Scenarios’ at RHIC” 72 (4): 1125–40. https://doi.org/10.1103/RevModPhys.72.1125.

Janzwood, Scott. 2021. “R&D priority-setting for global catastrophic risks: The case of the NASA planetary defense mission.” *Research Policy* 50 (6): 104225. https://doi.org/10.1016/j.respol.2021.104225

Janzwood, Scott. 2020. “Confident, Likely, or Both? The Implementation of the Uncertainty Language Framework in IPCC Special Reports.” *Climatic Change* 162 (3): 1655–75. https://doi.org/10.1007/s10584-020-02746-x.

Jasanoff, Sheila. 2004. “The Idiom of Co-Production.” In *States of Knowledge: The Co-Production of*

Science and Social Order, 1–12. New York: Routledge.

Johnson, Niall P. A. S., and Juergen Mueller. 2002. "Updating the Accounts: Global Mortality of the 1918-1920 Spanish Influenza Pandemic." *Bulletin of the History of Medicine* 76 (1): 105–15. https://doi.org/10.1353/bhm.2002.0022.

Jonassen, Rachael, and Roger Pielke. 2011. "Improving Conveyance of Uncertainties in the Findings of the IPCC." *Climatic Change* 108: 745–53. https://doi.org/10.1007/s10584-011-0185-7.

Jones, Roger N. 2011. "The Latest Iteration of IPCC Uncertainty Guidance: An Author Perspective." *Climatic Change* 108: 733–43. https://doi.org/10.1007/s10584-011-0239-x.

Joy, B. 2000. "Why the Future Doesn't Need Us." *Wired Magazine*, April 1, 2000. https://www.wired.com/2000/04/joy-2/.

Joyce, James M. 2005. "How Probabilities Reflect Evidence." *Philosophical Perspectives* 19 (Epistemology): 153–78.

Katzenstein, Peter. 1996. *The Culture of National Security: Norms and Identity in World Politics*. New York: Columbia University Press.

Kavadias, Stylianos, and Christoph H. Loch. 2004. *Project Selection Under Uncertainty: Dynamically Allocating Resources to Maximize Value*. Norwell, Massachusetts: Kluwer Academic Publishers.

Keisler, Jeffrey. 2004. "Value of Information in Portfolio Decision Analysis" 1 (3): 177–89.

Kennedy, J. V. 2012. "The Sources and Uses of U.S. Science Funding." *The New Atlantis* 36: 3–22.

Keur, P. van der, H. J. Henriksen, J. C. Refsgaard, M. Brugnach, C. Pahl-Wostl, A. Dewulf, and H. Buiteveld. 2008. "Identification of Major Sources of Uncertainty in Current IWRM Practice. Illustrated for the Rhine Basin." *Water Resources Management* 22 (11): 1677–1708. https://doi.org/10.1007/s11269-008-9248-6.

Keynes, John Maynard. 1921. *A Treatise on Probability*. London: Macmillan & Co.

King, David, Daniel Schrag, Zhou Dadi, Qi Ye, and Arunabha Ghosh. 2015. "Climate Change: A Risk Assessment." Cambridge, UK. http://www.csap.cam.ac.uk/media/uploads/files/1/climate-change--a-risk-assessment-v11.pdf.

Kleinmuntz, Don N. 2007. "Resource Allocation Decisions." In *Advances in Decision Analysis: From Foundations to Applications*, edited by Ward Edwards, Ralph F. Miles Jr., and Detlof von Winterfeldt, 400–418. Cambridge, UK: Cambridge University Press.

Knight, Frank H. 1921. *Risk, Uncertainty and Profit*. New York: Hart, Schaffner, and Marx.

Knol, Anne B., Arthur C. Petersen, Jeroen P. Van Der Sluijs, and Erik Lebret. 2009. "Dealing with Uncertainties in Environmental Burden of Disease Assessment." *Environmental Health: A Global*

Access Science Source 8 (1): 1–13. https://doi.org/10.1186/1476-069X-8-21.
Kopits, Elizabeth, Alex Marten, and Ann Wolverton. 2014. "Incorporating 'Catastrophic' Climate Change into Policy Analysis." *Climate Policy* 14 (5): 637–64. https://doi.org/10.1080/14693062.2014.864947.
Kowarsch, Martin, and Jason Jabbour. 2017. "Solution-Oriented Global Environmental Assessments : Opportunities and Challenges." *Environmental Science and Policy* 77: 187–92.
Kujala, Heini, Mark A. Burgman, and Atte Moilanen. 2013. "Treatment of Uncertainty in Conservation under Climate Change." *Conservation Letters* 6: 73–85. https://doi.org/10.1111/j.1755-263X.2012.00299.x.
Kwakkel, Jan H., Warren E. Walker, and Vincent A.W.J. Marchau. 2010. "Classifying and Communicating Uncertainties in Model-Based Policy Analysis." *International Journal of Technology, Policy and Management* 10 (4): 299–315. https://doi.org/10.1504/IJTPM.2010.036918.
Landis, Rob, and Lindley Johnson. 2019. "Advances in Planetary Defense in the United States." *Acta Astronautica* In Press: 1–15. https://doi.org/10.1016/j.actaastro.2018.06.020.
Langlois, R. N., and M. M. Cosgel. 1993. "Frank Knight on Risk, Uncertainty, and the Firm: A New Interpretation." *Economic Inquiry* 31: 456–65.
Lee, Robert C., Thomas D. Jones, and Clark R. Chapman. 2014. "A Decision Analysis Approach for Risk Management of Near-Earth Objects." *Acta Astronautica* 103: 362–69. https://doi.org/10.1016/j.actaastro.2013.12.014.
Lenton, T. M., H. Held, E. Kriegler, J. W. Hall, W. Lucht, S. Rahmstorf, and H. J. Schellnhuber. 2008. "Tipping Elements in the Earth's Climate System." *PNAS* 105 (6): 1786–93. http://www.pnas.org/cgi/doi/10.1073/pnas.0705414105.
Leslie, J. 1992. "Doomsday Revisited." *Philosophical Quarterly* 42 (166): 85–89.
Liu, Hin Yan, Kristian Cedervall Lauta, and Matthijs Michiel Maas. 2018. "Governing Boring Apocalypses: A New Typology of Existential Vulnerabilities and Exposures for Existential Risk Research." *Futures* 102: 6–19. https://doi.org/10.1016/j.futures.2018.04.009.
Ludwig, D. 1999. "Is It Meaningful to Estimate a Probability of Extinction?" *Ecology* 80 (1): 298–310.
MacAskill, William. 2015. *Doing Good Better: How Effective Altruism Can Help You Make a Difference*. New York: Avery.
Mach, Katharine J., Michael D. Mastrandrea, Patrick T. Freeman, and Christopher B. Field. 2017. "Unleashing Expert Judgment in Assessment." *Global Environmental Change* 44: 1–14. https://doi.org/10.1016/j.gloenvcha.2017.02.005.

Manheim, David. 2018. “Questioning Estimates of Natural Pandemic Risk.” *Health Security* 16 (6): 381–90. https://doi.org/10.1089/hs.2018.0039.

Martino, Joseph P. 1992. *Science Funding: Politics and Porkbarrel*. New Brunswick, New Jersey: Transaction Publishers.

———. 1995. *Research and Development Project Selection*. New York: John Wiley & Sons.

Mastrandrea, Michael D., Christopher B. Field, Thomas F. Stocker, Ottmar Edenhofer, Kristie L. Ebi, David J. Frame, Hermann Held, et al. 2010. “Guidance Note for Lead Authors of the IPCC Fifth Assessment Report on Consistent Treatment of Uncertainties.” *Intergovernmental Panel on Climate Change (IPCC)*. https://www.ipcc.ch/pdf/supporting-material/uncertainty-guidance-note.pdf.

Mastrandrea, Michael D., and Katharine J. Mach. 2011. “Treatment of Uncertainties in IPCC Assessment Reports: Past Approaches and Considerations for the Fifth Assessment Report.” *Climatic Change* 108: 659–73. https://doi.org/10.1007/s10584-011-0177-7.

Mastrandrea, Michael D., Katharine J. Mach, Gian Kasper Plattner, Ottmar Edenhofer, Thomas F. Stocker, Christopher B. Field, Kristie L. Ebi, and Patrick R. Matschoss. 2011. “The IPCC AR5 Guidance Note on Consistent Treatment of Uncertainties: A Common Approach across the Working Groups.” *Climatic Change* 108: 675–91. https://doi.org/10.1007/s10584-011-0178-6.

Matheny, Jason G. 2007. “Reducing the Risk of Human Extinction.” *Risk Analysis* 27 (5): 1335–44. https://doi.org/10.1111/j.1539-6924.2007.00960.x.

Matott, L. Shawn, Justin E. Babendreier, and S. Thomas Purucker. 2009. “Evaluating Uncertainty in Integrated Environmental Models: A Review of Concepts and Tools.” *Water Resources Research* 45 (6): 1–14. https://doi.org/10.1029/2008WR007301.

McMahon, Rosemarie, Michael Stauffacher, and Reto Knutti. 2015. “The Unseen Uncertainties in Climate Change: Reviewing Comprehension of an IPCC Scenario Graph.” *Climatic Change* 133: 141–54. https://doi.org/10.1007/s10584-015-1473-4.

Meah, Nafees. 2019. “Climate Uncertainty and Policy Making—What Do Policy Makers Want to Know?” *Regional Environmental Change*. https://doi.org/10.1007/s10113-019-01492-w.

Millett, Piers, and Andrew Snyder-Beattie. 2017. “Existential Risk and Cost-Effective Biosecurity.” *Health Security* 15 (4): 373–83. https://doi.org/10.1089/hs.2017.0028.

Mirakyan, Atom, and Roland De Guio. 2015. “Modelling and Uncertainties in Integrated Energy Planning.” *Renewable and Sustainable Energy Reviews* 46: 62–69. https://doi.org/10.1016/j.rser.2015.02.028.

Mishra, Harshit, Subhankar Karmakar, and Rakesh Kumar. 2018. “A Systematic Development of

Uncertainty Modeling in the Assessment of Health Risk to Environmental Contaminants." In *Environmental Contaminants: Measurement, Modelling and Control*, edited by Tarun Gupta, Avinash Kumar Agarwal, Rashmi Avinash Agarwal, and Nitin K. Labhsetwar, 195–207. Singapore: Springer Nature.

Morgan, M. Granger. 2014. "Use (and Abuse) of Expert Elicitation in Support of Decision Making for Public Policy." *PNAS* 111 (20): 7176–84. https://doi.org/10.1073/pnas.1319946111.

Morgan, M. Granger, and Max Henrion. 1990. *Uncertainty: A Guide to Dealing with Uncertainty in Quantitative Risk and Policy Analysis*. Cambridge, UK: Cambridge University Press.

Moss, R., and S Schneider. 2000. "Uncertainties in the IPCC TAR: Recommendations to Lead Authors for More Consistent Assessment and Reporting." In *Guidance Papers on the Cross Cutting Issues of the Third Assessment Report of the IPCC*, edited by T. Taniguchi and K. Tanaka R. Pachauri, 33--51. Geneva, Switzerland: World Meteorological Organization. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.399.6290&rep=rep1&type=pdf.

Mowery, David C. 2012. "Defense-Related R&D as a Model for 'Grand Challenges' Technology Policies." *Research Policy* 41: 1703–15. https://doi.org/10.1016/j.respol.2012.03.027.

Moynihan, Thomas. 2020. "Existential Risk and Human Extinction: An Intellectual History." *Futures* 116: 102495. https://doi.org/10.1016/j.futures.2019.102495.

Munthe, Christian. 2019. "The Black Hole Challenge: Precaution, Existential Risks and the Problem of Knowledge Gaps." *Ethics, Policy and Environment* 22 (1): 49–60. https://doi.org/10.1080/21550085.2019.1581415.

Murphy, C., Charles E. Harris, and P. Gardoni. 2011. "Classification and Moral Evaluation of Uncertainties in Engineering Modeling." *Science and Engineering Ethics* 17: 553–70. https://doi.org/10.1007/s11948-010-9242-2.

NASA. 2013. "Russia Meteor Not Linked to Asteroid Flyby." *NASA*, February 15, 2013. https://www.nasa.gov/mission_pages/asteroids/news/asteroid20130215.html.

———. 2018a. "FY 2019 Budget Estimates." https://www.nasa.gov/sites/default/files/atoms/files/fy19_nasa_budget_estimates.pdf.

———. 2018b. "Guidebook for Proposers Responding to a NASA Funding Announcement." https://www.hq.nasa.gov/office/procurement/nraguidebook/proposer2018.pdf.

———. 2018c. "ROSES-2018." https://nspires.nasaprs.com/external/viewrepositorydocument/cmdocumentid=611715/solicitationId=%7B22C6DBE5-E2E3-2089-5F2A-C01B0C2FF6EF%7D/viewSolicitationDocument=1/C.6 SSO

Lunar notice.pdf.

———. 2019a. "Mission Phase Definitions." NASA. 2019. https://science.nasa.gov/earth-science/programs/mission-phase-definitions.

———. 2019b. "PDCO Organization." 2019. https://www.nasa.gov/planetarydefense/organization.

National Research Council. 1995. "Research Prioritization: Analysis and Findings." In *Managing the Space Sciences*, 48–61. Washington, D.C.: The National Academies Press. https://doi.org/https://doi.org/10.17226/9297.

———. 2010. *Defending Planet Earth: Near-Earth-Object Surveys and Hazard Mitigation Strategies*. Washington, D.C.: The National Academies Press. https://doi.org/10.17226/12842.

Naumann, G., L. Alfieri, K. Wyser, L. Mentaschi, R. A. Betts, H. Carrao, J. Spinoni, J. Vogt, and L. Feyen. 2018. "Global Changes in Drought Conditions Under Different Levels of Warming." *Geophysical Research Letters* 45 (7): 3285–96. https://doi.org/10.1002/2017GL076521.

Near-Earth Object Science Definition Team. 2003. "Study to Determine the Feasibility of Extending the Search for Near-Earth Objects to Smaller Limiting Diameters." Washington, D.C.

———. 2017. "Update to Determine the Feasibility of Enhancing the Search and Characterization of NEOs." Washington, D.C.

Nearing, Grey S., Yudong Tian, Hoshin V. Gupta, Martyn P. Clark, Kenneth W. Harrison, and Steven V. Weijs. 2016. "A Philosophical Basis for Hydrological Uncertainty." *Hydrological Sciences Journal* 61 (9): 1666–78. https://doi.org/10.1080/02626667.2016.1183009.

Neff, Mark William. 2014. "Research Prioritization and the Potential Pitfall of Path Dependencies in Coral Reef Science." *Minerva* 52: 213–35. https://doi.org/10.1007/s11024-014-9250-5.

NOAA. 2020. "Weather Disasters and Costs." National Oceanic and Atmospheric Administration, Office for Coastal Management. 2020. https://coast.noaa.gov/states/fast-facts/weather-disasters.html.

Oppenheimer, Michael, Brian C O Neill, Mort Webster, and Shardul Agrawala. 2007. "The Limits of Consensus." *Science* 317 (September): 1505–7.

Oppenheimer, Michael, Brian C. O'Neill, and Mort Webster. 2008. "Negative Learning." *Climatic Change* 89: 155–72. https://doi.org/10.1007/s10584-008-9405-1.

Ord, T., R. Hillerbrand, and A. Sandberg. 2010. "Probing the Improbable: Methodological Challenges for Risks with Low Probabilities and High Stakes." *Journal of Risk Research* 132 (2): 191–205.

Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Oxford University Press.

Pelton, Joseph N. 2015. "International Cooperation and Collaboration in Planetary Defense Efforts." In *Handbook of Cosmic Hazards and Planetary Defense*, edited by Joseph N. Pelton and Firooz

Allahdadi, 1007–25. Springer International Publishing. https://doi.org/10.1007/978-3-319-03952-7_74.

Pereira, Joana Castro, and Eduardo Viola. 2018. “Catastrophic Climate Change and Forest Tipping Points: Blind Spots in International Politics and Policy.” *Global Policy* 9 (4): 513–24. https://doi.org/10.1111/1758-5899.12578.

Petersen, Arthur C. 2012. *Simulating Nature: A Philosophical Study of Computer-Simulation Uncertainties and Their Role in Climate Science and Policy Advice*. Second. Boca Raton, Florida: CRC Press.

Petersen, Arthur C., Peter H. M. Janssen, Jeroen P. van der Sluijs, and James S. Risbey and Jerome R. Ravetz. 2003. “RIVM/MNP Guidance for Uncertainty Assessment and Communication: Mini-Checklist & Quickscan Questionnaire.” Bilthoven, The Netherlands. https://www.rivm.nl/bibliotheek/digitaaldepot/Guidance_MC_QS-Q.pdf.

Pindyck, Robert S. 2011. “Fat Tails, Thin Tails, and Climate Change Policy.” *Review of Environmental Economics and Policy* 5 (2): 258–74. https://doi.org/10.1093/reep/rer005.

Popper, K. R. 1982. *The Open Universe: An Argument for Indeterminism*. London: Hutchinson.

Porter, T. M. 1995. *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton University Press.

Rabinovich, Anna, and Thomas A. Morton. 2012. “Unquestioned Answers or Unanswered Questions: Beliefs About Science Guide Responses to Uncertainty in Climate Change Risk Communication.” *Risk Analysis* 32 (6): 992–1002. https://doi.org/10.1111/j.1539-6924.2012.01771.x.

Raftery, Adrian E, Alec Zimmer, Dargan M W Frierson, Richard Startz, and Peiran Liu. 2017. “Less than 2 ◦ C Warming by 2100 Unlikely.” *Nature Climate Change* 7 (September 2017): 637–42.

Raiffa, H., and R. O. Schlaifer. 1961. *Applied Statistical Decision Theory*. Cambridge, Massachusetts: Division of Research, Graduate School of Business Administration, Harvard University.

Ramsey, Frank P. 1926. “Truth and Probability.” In *The Foundations of Mathematics and Other Logical Essays*, edited by R. B. Braithwaite. London: Routledge and Kegan Paul.

Raup, D. M., and J. J. Sepkoski. 1982. “Mass Extinctions of the Marine Fossil Record.” *Science* 215 (4539): 1501–3.

Rees, Martin. 2013. “Denial of Catastrophic Risks.” *Science* 339: 1123.

Refsgaard, Jens Christian, K. Arnbjerg-Nielsen, M. Drews, K. Halsnæs, E. Jeppesen, H. Madsen, A. Markandya, J. E. Olesen, J. R. Porter, and J. H. Christensen. 2013. “The Role of Uncertainty in Climate Change Adaptation Strategies-A Danish Water Management Example.” *Mitigation and*

Adaptation Strategies for Global Change 18 (3): 337–59. https://doi.org/10.1007/s11027-012-9366-6.

Refsgaard, Jens Christian, Jeroen P. van der Sluijs, Anker Lajer Højberg, and Peter A. Vanrolleghem. 2007. "Uncertainty in the Environmental Modelling Process: A Framework and Guidance." *Environmental Modelling and Software* 22 (11): 1543–56. https://doi.org/10.1016/j.envsoft.2007.02.004.

Regan, Helen M, Mark Colyvan, and Mark A Burgman. 2002. "A Taxonomy and Treatment of Uncertainty for Ecology and Conservation Biology." *Ecological Applications* 12 (2): 618–28.

Reich, Eugenie Samuel. 2010. "NASA Panel Weighs Asteroid Danger: Telescopes in Space Could Help Pin down the Risk of a Deadly Impact." *Nature* 467 (7312): 140–41.

Renn, Ortwin. 2008. *Risk Governance: Coping With Risk in a Complex World*. Sterling, VA: Earthscan.

Rittel, Horst W. J., and Melvin M. Webber. 1973. "Dilemmas in a General Theory of Planning." *Policy Sciences* 4: 155–69.

Robinson, Douglas K.R., and Mariana Mazzucato. 2019. "The Evolution of Mission-Oriented Policies: Exploring Changing Market Creating Policies in the US and European Space Sector." *Research Policy* 48 (4): 936–48. https://doi.org/10.1016/j.respol.2018.10.005.

Royde-Smith, John Graham, and Thomas A. Hughes. 2019. "World War II." *Encyclopædia Britannica*. https://www.britannica.com/event/World-War-II/Costs-of-the-war#ref53608.

Ruggie, J. G. 1998. "What Makes the World Hang Together? Neo-Utilitarianism and the Social Constructivist Challenge." *International Organization* 52 (4): 855–85.

———. 2014. "Global Governance and 'New Governance Theory': Lessons from Business and Human Rights." *Global Governance* 20: 5–17.

Runde, J. 1998. "Clarifying Frank Knight's Discussion of the Meaning of Risk and Uncertainty." *Cambridge Journal of Economics* 22: 539-546.

Sandberg, Anders. 2017. "Societal and Ethical Issues Related to Catastrophic and Existential Risk." In *Proceedings of the First International Colloquium on Catastrophic and Existential Risk*, 107–24. Los Angeles: The B. John Garrick Institute for the Risk Sciences, University of California Los Angeles. https://www.risksciences.ucla.edu/news-events/2018/1/2/proceedings-of-the-first-international-colloquium-on-catastrophic-and-existential-risk.

Sandberg, Anders, and Nick Bostrom. 2008. "Global Catastrophic Risks Survey." Technical Report #2008-1. Oxford, UK. https://www.fhi.ox.ac.uk/reports/2008-1.pdf.

Sigel, Katja, Bernd Klauer, and Claudia Pahl-Wostl. 2010. "Conceptualising Uncertainty in

Environmental Decision-Making: The Example of the EU Water Framework Directive." *Ecological Economics* 69 (3): 502–10. https://doi.org/10.1016/j.ecolecon.2009.11.012.

Skeels, Meredith, Bongshin Lee, Greg Smith, and George G. Robertson. 2010. "Revealing Uncertainty for Information Visualization." *Information Visualization* 9 (1): 70–81. https://doi.org/10.1057/ivs.2009.1.

Skinner, Daniel J.C., Sophie A. Rocks, Simon J.T. Pollard, and Gillian H. Drew. 2014. "Identifying Uncertainty in Environmental Risk Assessments: The Development of a Novel Typology and Its Implications for Risk Characterization." *Human and Ecological Risk Assessment* 20 (3): 607–40. https://doi.org/10.1080/10807039.2013.779899.

Soros, George. 2013. "Fallibility, Reflexivity, and the Human Uncertainty Principle." *Journal of Economic Methodology* 20 (4): 309–29. https://doi.org/10.1080/1350178X.2013.859415.

Spiegelhalter, David J., and Hauke Riesch. 2011. "Don't Know, Can't Know: Embracing Deeper Uncertainties When Analysing Risks." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 369 (1956): 4730–50. https://doi.org/10.1098/rsta.2011.0163.

Spratt, David, and Ian Dunlop. 2018. *What Lies Beneath: The Understatement of Existential Climate Risk*. Melbourne, Australia: Breakthrough: National Centre for Climate Restoration.

Steffen, Will, Johan Rockström, Katherine Richardson, Timothy M Lenton, Carl Folke, Diana Liverman, Colin P Summerhayes, et al. 2018. "Trajectories of the Earth System in the Anthropocene." *PNAS* 115 (33): 8252–59. https://doi.org/10.1073/pnas.1810141115/-/DCSupplemental.

Stewart, Jenny. 1995. "Models of Priority-Setting for Public Sector Research." *Research Policy* 24: 115–26. https://doi.org/10.1016/0048-7333(93)00752-F.

Stirling, Andrew. 1998. "Risk at a Turning Point?" *Journal of Risk Research* 1 (2): 97–109.

Stone, Deborah. 1989. "Causal Stories and the Formation of Policy Agendas." *Political Science Quarterly* 104 (2): 281–300.

Svozil, K. 1996. "Undecidability Everywhere?" In *Boundaries and Barriers: On the Limits to Scientific Knowledge*, edited by J.L. Casti and A. Karlqvist, 215–37. Reading, Massachusetts: Addison-Wesley.

Swart, Rob, Lenny Bernstein, Minh Ha-Duong, and Arthur Petersen. 2009. "Agreeing to Disagree: Uncertainty Management in Assessing Climate Change, Impacts and Responses by the IPCC." *Climatic Change* 92: 1–29. https://doi.org/10.1007/s10584-008-9444-7.

Taleb, N. N. 2007. *The Black Swan: The Impact of the Highly Improbable*. New York: Random House.

Tegmark, Max, and Nick Bostrom. 2005. “Is a Doomsday Catastrophe Likely?” *Nature* 438 (7069): 754. https://doi.org/10.1038/438754a.

Tonn, Bruce, and Dorian Stiefel. 2013. “Evaluating Methods for Estimating Existential Risks.” *Risk Analysis* 33 (10): 1772–87. https://doi.org/10.1111/risa.12039.

Torres, Phil. 2015. “Problems with Defining an Existential Risk.” *Institute for Ethics and Emerging Technologies*, January 21, 2015. https://ieet.org/index.php/IEET2/more/torres20150121.

Tsur, Yacov, and Cees Withagen. 2013. “Preparing for Catastrophic Climate Change.” *Journal of Economics* 110: 225–39. https://doi.org/10.1007/s00712-012-0331-3.

Turchin, Alexey, and David Denkenberger. 2018. “Global Catastrophic and Existential Risks Communication Scale.” *Futures* 102: 27–38. https://doi.org/10.1016/j.futures.2018.01.003.

Turco, R. P., O. B. Toon, T. P. Ackerman, J. B. Pollack, and C. Sagan. 1983. “Nuclear Winter: Global Consequences of Multiple Nuclear Explosions.” *Science* 222 (4630): 1283–89.

Vesely, W. E., and D. M. Rasmuson. 1984. “Uncertainties in Nuclear Probabilistic Risk Analyses.” *Risk Analysis* 4: 313–322.

Walker, W. E., P. Harremoës, J Rotmans, J. P van der Sluijs, M B A van Asselt, P Janssen, and M P Krayer von Krauss. 2003. “Defining Uncertainty: A Conceptual Basis for Uncertainty Management in Model-Based Decision Support.” *Integrated Assessment* 4 (1): 5–17. https://doi.org/10.1076/iaij.4.1.5.16466.

Wallace, Matthew L., and Ismael Ràfols. 2015. “Research Portfolio Analysis in Science Policy: Moving from Financial Returns to Societal Benefits.” *Minerva* 53: 89–115. https://doi.org/10.1007/s11024-015-9271-8.

———. 2018. “Institutional Shaping of Research Priorities: A Case Study on Avian Influenza.” *Research Policy* 47 (10): 1975–89. https://doi.org/10.1016/j.respol.2018.07.005.

Watts, Nick, Markus Amann, Sonja Ayeb-Karlsson, Kristine Belesova, Timothy Bouley, Maxwell Boykoff, Peter Byass, et al. 2018. “The Lancet Countdown on Health and Climate Change: From 25 Years of Inaction to a Global Transformation for Public Health.” *Lancet* 391: 581–630. https://doi.org/10.1016/S0140-6736(17)32464-9.

Weitzman, Martin L. 2008. “On Modeling and Interpreting the Economics of Catastrophic Climate Change.” *The Review of Economics and Statistics* 91 (1): 1–19.

Winsberg, Eric. 2018. *Philosophy and Climate Science*. Cambridge: Cambridge University Press.

Winterfeldt, Detlof von, and Ward Edwards. 2007. *Defining a Decision Analytic Structure*. Edited by Ward Edwards, Ralph F. Miles Jr., and Detlof von Winterfeldt. *Advances in Decision Analysis:*

From Foundations to Applications. Cambridge, UK. https://doi.org/10.1017/CBO9780511611308.007.

Woodward, James. 1998. “Statistics.” In *Routledge Encyclopedia of Philosophy*. Taylor and Francis. https://doi.org/10.4324/9780415249126-Q101-1.

World Bank. 2012. *Turn Down the Heat: Why a 4°C Warmer World Must Be Avoided*. New York: World Bank.

Wüthrich, Nicolas. 2017. “Conceptualizing Uncertainty: An Assessment of the Uncertainty Framework of the Intergovernmental Panel on Climate Change.” In *EPSA15 Selected Papers: The 5th Conference of the European Philosophy of Science Association in Düsseldorf*, edited by Michela Massimi, Jan-Willem Romeijn, and Gerhard Schurz, 95–107. Springer.

Wynne, Brian. 1992. “Uncertainty and Environmental Learning. Reconceiving Science and Policy in the Preventive Paradigm.” *Global Environmental Change* 2 (2): 111–27. https://doi.org/10.1016/0959-3780(92)90017-2.

Xu, Yangyang, and Veerabhadran Ramanathan. 2017. “Well below 2 °C: Mitigation Strategies for Avoiding Dangerous to Catastrophic Climate Changes.” *PNAS* 114 (39): 10315–10323. https://doi.org/10.1073/pnas.1618481114.

Yokota, Fumie, and Kimberly M. Thompson. 2004. “Value of Information Literature Analysis: A Review of Applications in Health Risk Management.” *Medical Decision Making* 24 (3): 287–98. https://doi.org/10.1177/0272989X04263157.

Appendix A

Typologies, taxonomies, and categorizations of uncertainty not included in Skinner et al. (2014)

Author(s)	Uncertainty categories
Faber, Manstetten, & Proops (1992)	[Types of “surprise”] risk; uncertainty; ignorance [Ignorance] reducible ignorance; irreducible ignorance
Baecher & Christian (2000)	[Type]: natural variation; lack of knowledge
Chen, Ma, & Reckhow (2007)	[Type]: qualitative uncertainty; quantitative uncertainty
Refsgaard et al. (2007)	[Nature]: epistemic; stochastic [Level]: statistical; scenario; qualitative; recognized ignorance [Location]: context; inputs; model; model outputs
van der Keur et al. (2008)	[Level]: statistical; scenario; qualitative; recognized ignorance; total ignorance
Brugnach et al. (2008)	[Types]: unpredictability; incomplete knowledge; multiple knowledge frames
Matott, Babendreier, and Purucker (2009)	[Types]: purely irreducible; partly (ir)reducible; purely reducible; certain [Sources]: qualitative modeler uncertainty; quantitative input uncertainty
Briggs, Sabel, and Lee (2009)	[Nature]: intrinsic; extrinsic [Location]: conceptual; analytical; communicational; [Magnitude]: statistical; scenario; recognized ignorance
Skeels et al. (2010)	[Types]: credibility; disagreement [Levels]: measurement precision; completeness; inferences

Sigel et al. (2010)	[Types]: fact-related; norm-related [Causes]: Phenomenological; epistemological [Levels]: level of knowledge; level of confidence about knowledge
Dequech (2011)	[Types]: weak; strong (ambiguity, fundamental)
Murphy, Harris, & Gardoni (2011)	[Developing a model]: model inexactness; mistaken assumptions; measurement error; statistical uncertainty [Applying or implementing a model]: randomness in the variables; volitional uncertainty; human error [Results of a model]: endoxastic uncertainty; metadoxastic uncertainty [Nature]: aleatory; epistemic
Spiegelhalter and Riesch (2011)	[Levels]: uncertainty about which in a list of events will occur; parameter uncertainty; uncertainty about which model is best; uncertainty about known inadequacies of best model; uncertainty about unknown inadequacies of all models
Petersen (2012)	[Categories]: location; nature; range; recognized ignorance; methodological unreliability; value diversity
Bradley & Drechsler (2013)	[Types]: ethical; option; state space; state; empirical [Nature]: modal; empirical; normative [Object]: factual; counterfactual [Severity]: ignorance; severe uncertainty; mild uncertainty; certainty
Kujala et al. (2013)	[Types]: linguistic uncertainty; human decision uncertainty; epistemic uncertainty
Refsgaard et al. (2013)	[Nature]: aleatory; epistemic; ambiguity [Level]: statistical; scenario; qualitative; recognized ignorance; total ignorance [Source] input data; model uncertainty; context uncertainty; uncertainty due to multiple knowledge frames
Skinner et al. (2014)	[Nature]: aleatory; epistemic; combined [Level]: deterministic; statistical; scenario; recognized ignorance; total ignorance [Location]:

	data, language, system, variability, extrapolation, model, decision
Mirakyan & De Guio (2015)	[Categories]: linguistic, knowledge (epistemic), variability (aleatoric), decision, level, procedural [Level]: ignorance, uncertainty, risk
Beven (2016)	[Types]: aleatory; epistemic (system dynamics, forcing and response data, disinformation); semantic/linguistic; ontological
Nearing et al. (2016)	[Types]: uncertainty related to our ability to simulate the universe; uncertainty related to our knowledge about phenomena that occur in the universe
Mishra et al. (2018)	[Location]: data; linguistic; system; variability; extrapolation; model; decision [Level]: deterministic; statistical; scenario; recognized ignorance; total ignorance [Nature]: aleatory; epistemic; combined

Appendix B

Alternative interpretations of the IPCC's confidence and likelihood scales

The following addendum was added as "Supplementary Material" to Janzwood 2020.

Aven offers a slightly different interpretation of the IPCC's confidence and likelihood scales, based on the Bayesian interpretation of probability as "the analyst's uncertainty or degree of belief that … 'statement A' is true" (2019, pp. 8). He argues that neither the confidence nor likelihood scales constitute clear statements of probability. Aven sees the confidence scale as a statement of the "strength of knowledge" supporting an implied probability distribution, while the likelihood terms (largely based on frequencies) merely reflect "statistically expected values" – not probabilities, as he defines them. As a result, Aven recommends that frequentist probabilities be clearly identified and incorporated into a (subjective) probability assessment reflecting the analyst's degree belief in the "truthfulness" of the knowledge claim. Confidence terms should then be understood as an assessment of the strength of knowledge supporting those probabilities.

However, in this paper, I do not make this added distinction between probabilities and strength of knowledge assessments. Instead, my analysis reflects the position that strength of knowledge assessments can be embedded within subjective probabilities and therefore, confidence statements are themselves (qualitative) statements of subjective probability. The position that subjective probabilities can be updated to reflect the analyst's assessment of the strength of their knowledge is championed by Howard who states: "Once you have internalized the thinking of Laplace and Jaynes, any notion of uncertainty about probability becomes unnecessary" (Howard 2007, pp. 50).