Addressing Risk Governance Deficits through Scenario Modeling Practices

by
John Benjamin Cassel

Submitted to OCAD University in partial fulfillment of the
requirements for the degree of
Master of Design
in
Strategic Foresight and Innovation

Toronto, Ontario, Canada, August 2011

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

I authorize OCAD University to lend this MRP to other institutions or individuals for the purpose of scholarly research.

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

I further authorize OCAD University to reproduce this MRP by photocopying or by other means, in total or in part, at the request of other institutions or individuals for the purpose of scholarly research.

Signature ______________________________

Abstract

In a world of inevitable regret, those governing risk must build practices that withstand the vicissitudes of actual events by demonstrating that reasonable efforts had been and will continue to be taken despite those harms. However, what is reasonable depends on one's worldview, and so not giving different worldviews appropriate consideration leads to deficits in the quality of risk governance. This project developed foresight methods for eliciting, discovering, representing, and modeling scenarios which capture the counterfactual forests created by disparate worldviews. These methods employ structural differences between objective and subjective relations toward physical events to delineate the actual points of contention, while maintaining neutrality by remaining strictly grounded in the input of the stakeholders themselves. These methods respect how people frame causal information psychologically, avoiding biases known to affect political judgment. Overall, these methods serve as a reminder that how we ask designs how we think.

Acknowledgments

My kindest thanks to Peter Jones and Walter Derzko, who ably advised and enthusiastically supported this project from its origin as a class project, as taught by Peter and Suzanne Stein. Nabil Harfoush also ably served as an adviser early on in the project.

Thanks to my collaborators on that class project, Karl Schroeder and Andrew McKee. Karl helped me extend the work proposed there, by undergoing a prototype interview and by instigating my search for how to evaluate discovery processes. Thanks to all of my other classmates and teachers in the first Strategic Foresight and Innovation program, among which I thank Doug Reid, Susan Gorbet, Jon Resnick, Anne Sutherland, and Peter Rose specifically for various insights used here.

My kindest thanks to my wife Sarah Mongeau for being the best of our adventures. I thank all of my family and friends, as they have experienced the joys and sorrows of life largely without me for far too long. I specifically thank Wayne Davis, Duane Searsmith, and Michael Hamman for their ongoing support. My thanks to Wolfram|Alpha for continuing to employ me part-time throughout.

Finally, I wish to thank the scholars and writers who loom so large in my mind through their words and work. These include Philip Tetlock, Joshua Tenenbaum, Judea Pearl, Alison Gopnik, Vaclav Smil, Cosma Shalizi, Ortwin Renn, Charles Yu, Manuel de Landa, Jamais Cascio, and many more.

Contents

List of Tables

List of Figures

The Opportunity of a Technical Foresight for Risk Governance

Foresight: the Inverse of History

> *"I cannot take architecture out of thin air. . . but I certainly enjoy taking it out of thick air!"*
>
> -from *"The City, Seen as a Garden of Ideas"* by Peter Cook [Cook, 2003]

History has a special privilege that is not shared by most subjects. This privilege is a freedom to pay attention to any topic, in any form, with any focus, as long as that investigation yields an academically firm insight about the past. This freedom is appropriate, as our histories must be as rich, diverse, deep, and *thick with context* as the varied cultures, ideas, peoples, economies, and practices that inhabited that past. If there is a field that is mandated to transcend disciplines, and further, to build the methods of transcending discipline, it is history.

There is a subject that shares this freedom (and the concomitant responsibilities), and that is its inverse, the study of possible futures and how they were, are, and will come to be considered. This discipline is called foresight. It too studies every corner of the future, from the visions put forth in popular culture, to feverish odd ideas bubbling within science fiction, to every kind of prediction and projection, to all coordination and institutional policies and arrangements, and to the everyday plans of individuals and how they go wrong.

It is within this context that I recognize that doing work in foresight brings a special responsibility, to put forth an academic lens appropriate for examining of the hopes, dreams, fears, concerns, and uncertainties of our combined futures. In short, good works of foresight present futures worthy of building a conceptual present in the same way good histories builds us suitable pasts.

While I have little hope of accomplishing this, I can at least pick the right materials. A great history will tell us of how a people attempted to come together to work through a terrible threat, for which the outcomes were uncertain, and how they, despite inconsistencies, false starts, and setbacks, were changed by those events, yielding either a story of success worthy for emulation, or a story of woe and regrets producing warnings and caution. It is for this reason that I focus on the risks of our time that seem to threaten the most catastrophic consequences, how those risks are conceived, and the processes at work by which the conception of those risks could serve our human ends.

The Distributed Risk of Infrastructure and its Transitions

The 2010 volcanic eruptions of Eyjafjallajökull in Iceland disrupted the travel of ten million passengers, leading to losses estimated between 1.5 and 2.5 billion Euros [Rincon, 2011]. How is it that these losses came about, with so many companies unprepared for that contingency? It is not because there were never eruptions in Iceland before, and it also is not because nobody imagined that Icelandic volcanoes might erupt again. It is because volcanic eruptions of that nature are rare, have nothing to do directly with any of the business or personal motivations that drove people to take the flights in the first place, and are entirely divorced from the routine of everyday life. Volcanic eruptions are just one of thousands of random possibilities that one effectively cannot think about if one is to get anything done.

Fortunately, in the case of the 2010 Eyjafjallajökull eruption, we as a civilization have largely learned to compensate for these disasters. Although the disruption in business was costly, and possibly could have been prevented to some degree, the risk mitigation measures of shutting down the airspace prevented any actual flight disasters. Though stranded in foreign countries, travelers did manage to find provisions. Altogether, institutions for understanding, monitoring, and mitigating the risk of specific events are in place, although the tools to effectively spread the knowledge of these contingencies to a broader risk-management audience may yet be limited. Indeed, it is more common to rely upon indemnifying these risks through business continuity insurance, or more common yet, given the assumption of an supporting institutional backdrop, simply enduring them.

However, although we have set up institutions that arbitrate direct infrastructural risks in the environments of our daily lives well enough for many, it is fair to say that we have not figured out how to make sense of the risks posed by transitions in this infrastructure. Everyday life entails a vast web of materials, processing, energy production, and land use, but it is impossible to weigh our daily activities against the ongoing and eventual consequences in any sensible way, especially given the diversity of existing stakes, interests, perceptions, structures, processes, and biases involved in those subjects. Most challenging are issues that are highly distributed in time, space, and context; pitting alternatives with uncertain rewards and heavy short-term individual costs under specific regulatory frameworks against scientifically-uncertain irreversible long-term public costs that span borders (which we will refer to as *distributed risk challenges*). It has yet to be demonstrated if our industrial infrastructure (and its supporting institutions) can successfully make large-scale transitions in response to a changing portfolio of energy, material, and ecosystem service availability.

At the time of this writing, the most salient catastrophe is the March 2011 earthquakes and tsunamis striking Japan, and the ensuing issues in controlling the Fukushima nuclear power plant. One may question why an island along the fault-lines of tectonic plates would then turn to nuclear power, given its dangers. Yet, nuclear power, despite the direct and long-term risks posed by radiation and reaction, is thought by some to be a comparatively progressive technology, without the climate risks many attribute to carbon-based energy technologies [Cravens, 2007] [Brand, 2009]. Would Japan's nuclear infrastructure have changed the anyone's opinion of the overall acceptability to nuclear power if not for this catastrophe? Is it appropriate that this catastrophe might disrupt a progressive infrastructure of energy provision?

The Imponderable Character of Irreversibility and Regret

> *"...trying to come to terms with how screwed up and unfair it is that we only get to do this all once, with the intractability and general awfulness of trying to parse the idea of once, trying to get any kind of handle on it,... the slippery idea of onceness."*
> -from *"How to Live Safely In a Science Fictional Universe"* by Charles Yu [Yu, 2010]

Regret is the disparity between what did happen under what actions were undertaken and what would be projected to happen under a different course of action. Following from the previous section, we have many regrets that their infrastructure was not engineered differently. However, regrets are curious things, for regrets are relative, not absolute, harms. For there to be a regret, there must be another state-of-affairs that did not occur, but that we take so seriously as a possibility that we judge ourselves in reference to it [Gopnik, 2009]. These other states-of-affairs are counter to the facts of what happened, and thus are called counterfactuals. As there are no future facts, but there are many future states of affairs that could be. Therefore, it is appropriate that this study of possible futures, or foresight, should concern itself with counterfactuals, how they are understood, and how they are applied.

We would like to avoid regrets, but is impossible. Imagine that you are walking down the street and are asked for spare change. Whether or not we comply with the request, we experience regret, at either being taken advantage of and deprived of what is ours, or for not showing mercy to fellow person with sufficient needs as to make that demand. It is for this reason that there is no such thing as perfect precaution: we are, and will be, in circumstances that require conflicting trade-offs between value choices that have no objective answer.

Risk considered at broader scales does not escape the ambiguity of mutual regret. When taken to its logical extreme, the Precautionary Principle, which is the guiding principle of European regulatory philosophy, and advises to mitigate against costly risks even when the consequences cannot be scientifically assured, turns from wise council to a recipe for paralysis [Sunstein, 2007]. The provision of infrastructure will always have consequences, and change will always have the particular risks associated with the untried, untested, and experimental. To mitigate against any action which might be risky would be to do nothing, except for the extraordinary risk that doing nothing entails.

The bitterest regrets are those that are irreversible. An injured individual can be treated and compensated, and an endangered species can be given special protections, but we can give nothing to the dead and extinct. When facing infrastructural trade-offs with irreversible consequences, how can we even conceive of right choices, much less how to see that those choices get made?

Severe Risks and their Governance

Given the imponderable nature of risk and regret, it is essential to have a stronger response than designing approaches assured of success. These approaches are far too limited given the portfolio of regrets we face. A better way to meet that challenge is to build governing approaches, approaches that will be recognized by those for whom it fails that it was, nonetheless, a reasonable option given what was known and what consensuses were achieved. The challenge is discovering options that we can all recognize, given our limits, make the best trade-offs. In short, given that we cannot avoid risk, our task will be to see that we have done due diligence in avoiding and mitigating it to the degree possible.

This focus on risk does little to reduce the domain of foresight. There are few concepts as trans-disciplinary as the future, but risk is one of them; consider the disciplinary breadth to which risk is considered seriously: economics, law, sociology, psychology, political science, applied mathematics, statistics, artificial intelligence, epidemiology, philosophy, insurance, and finance. It is within this scope that we discover the field of risk governance [Renn, 2008]. Risk governance combines such divergent approaches such as post-modern social criticism, engineering reliability studies, systems theory, and insurance into a pragmatic, Habermassian framework containing the best of both the analytical and deliberative worlds.

Risk governance provides two distinctive advantages as a framework. First of all, it captures a high-level methodological preference, which is to favor a certain pluralism or multiplicity, instead of unification, as the most appropriate way to frame problems at their broadest sense. This captures the trend toward governance generally, where *governance embodies a horizontally organized structure of functional self-regulation encompassing state and non-state actors bringing about collectively binding decisions without superior authority* [Rosenau, 1992] [Wolf, 2002] [Wolf, 2005]. This move toward accommodating fragmentation may ultimately be a personal preference, and more adept designers may be able to develop unified systems. Yet, this fragmentation is something culturally real, as we perceive ourselves not as a participant of a single system, but many, serving different roles within them.

This distinction will also reappear in terms of design methodology used here, where we will see a clear preference for approaches used to make sense of one's current context over those that strictly define or delimit systems, in resonance with the multi-ontology perspective that is demanded when making useful sense of our world [Snowden, 2005].

Further, this ability to engage multiplicity extends to embracing both sides of long-standing distinctions and differences, such as evidence/values, objective/subjective, social constructivism/realism, qualitative/quantitative, and so forth. If nothing else, the respect to these distinctions is pragmatic: the effectiveness of the risk governance depends on appropriately engaging stakeholders with different views on these subjects.

The second core advantage of working within the risk governance framework is that it offers us standards of due diligence. What does appropriate due diligence consist of? One possible answer is that it means consistently and competently applying a variety of common-sense measures known to eliminate common pitfalls. Fortunately,

Table 1: Governance Deficits in Assessing and Understanding Risks

A1: The failure to detect early warnings of risk because of erroneous signals, misinterpretation of information, or simply not enough information being gathered

A2: The lack of adequate factual knowledge for robust risk assessment because of existing gaps in scientific knowledge or failure to either source existing information or appreciate its associated uncertainty

A3 : The omission of knowledge related to stakeholder risk perceptions and concerns

A4: The failure to consult the relevant stakeholders, as their involvement can improve the information input and the legitimacy of the risk assessment process (provided that interests and bias are carefully managed)

A5: The failure to properly evaluate a risk as being acceptable or unacceptable to society

A6: The misrepresentation of information about risk, whereby biased, selective or incomplete knowledge is used during, or communicated after, risk assessment, either with or without intention

A7: A failure to understand how the components of a complex system interact or how the system behaves as a whole, thus a failure to assess the multiple dimensions of a risk and its potential consequences

A8: A failure to recognize fast or fundamental changes to a system, which can cause new risks to emerge or old ones to change

A9: The inappropriate use of formal models as a way to create and understand knowledge about complex systems (over- and under-reliance on models can be equally problematic)

A10: A failure to overcome cognitive barriers to imagining that events outside expected paradigms are possible

Table 2: Governance Deficits in Managing Risks

B1: A failure to respond adequately to early warnings of risk, which could mean either under or over-reacting to warnings

B2: A failure to design effective risk management strategies. Such failure may result from objectives, tools, or implementation plans being ill-defined or absent

B3: A failure to consider all reasonable, available options before deciding how to proceed

B4: Not conducting appropriate to assess the costs and benefits (efficiency) of various options and how these are distributed (equity)

B5: A failure to implement risk management strategies or policies and to enforce them

B6: A failure to anticipate the consequences, particularly negative side effects, of a risk management decision, and to adequately monitor and react to the outcomes

B7: An inability to reconcile the time-frame of the risk issue (which may have far-off consequences and require a long-term perspective) with decision-making pressures and incentives (which may prioritize visible, short-term results or costly reductions)

B8: A failure to adequately balance transparency and confidentiality during the decision-making process, which can have implications for stakeholder trust or for security

B9: A lack of adequate organizational capacity (assets, skills, and capabilities) and/or of a suitable culture (one that recognizes the value of risk management) for ensuring managerial effectiveness when dealing with risks

B10: A failure of the multiple departments or organizations responsible for a risk's management to act individually but cohesively, or of one entity to deal with several risks

B11: A failure to deal with the complex nature of commons problems, resulting in inappropriate or inadequate decisions to mitigate commons-related risks (e.g. risks to the atmosphere or oceans)

B12: A failure to resolve conflicts where different pathways to resolution may be required in consideration of the nature of the conflict and of different stakeholder interests and values

B13: Insufficient flexibility or capacity to respond adequately to unexpected events because of bad planning, inflexible mindsets, and response structures, or an inability to think creatively and innovate when necessary the risk governance community has developed lists of such common pitfalls, which they refer to as risk governance deficits [Graham et al., 2009], including a list of deficits in assessing and understanding risk (see Table 1) and a list for deficits in managing risk (see Table 2). These lists are hard won, accumulating the reflections of what went wrong in wide variety of risk issues, and as best as can reasonably be determined constitute wise council. Furthermore, these deficits touch on a wide variety of general considerations, including knowledge management, knowledge of stakeholder benefits and harms, perceptions, preparation, timing, modeling, organizational structures, and other issues. Therefore, it is a reasonable working assumption that these guidelines, even if not comprehensive, are sufficient for building excellent risk governance methods and tools.

Worldview as an Obstacle to Governing Risk

Policy problems can be made difficult by both their severity, their complexity, and their ambiguity. A common-sense measure for severity of many policy problems is the combined severity of the impact on those whom the policy will intervene. However, those impacted may have such different worldviews that it is challenging to determine exactly how these impacts function. One particular challenge is when different concerns are in conflict, but the worldviews guiding those are so intractably different that the exact conflicts are difficult to communicate appropriately.

Strikingly, problems with some of the most severe consequences (death, injury, property damage, and environmental degradation) are physical in character, and can be connected to physical causes. By separating these physical causes and the empirical processes that govern them from the impacts as reported after these consequences have been established, we formulate a useful barrier: those in which we can readily admit other observations and explanations that contradict established evidence (i.e. the objective), from those for which the unique rights of determination emerge from the coherent reaction of the experiencier (i.e. the subjective). For example, medical evidence cannot determine that your leg hurts, but it can determine that it is broken. It is through this distinction that distributed risk problems, although severe in their outcomes, seem more addressable than certain cultural and social policy issues, by virtue of their widely-recognized objective benefits and harms.

Unfortunately, the use of these distinctions is necessary, as high-impact problems pose different kinds of conflicts between stakeholders. The stakeholders may have different subjective views: i.e. if they were to come to a consensus on an objective, material comparison of situations (and further, agreement that these situations are objective and material), they will not come to the same normative conclusions as to the more significant aspects of these situations, and thus have no consensus on which is desirable. Yet, at the same time, it is entirely unclear to the degree that participants share a common objective picture. They may interpret the same event as implying that different underlying processes were taking place. Further, their assessment of the normative judgment of other stakeholders may be suspect. To compound all of these factors, the stakeholders may not differentiate between these kinds of assessments and misassessments.

An appropriate method for clarifying and arranging these distinctions should imply and justify potentially unexpected consequences in the limited cases in which it is unambiguously successful, and draw critical attention to structural problems when they may undermine other analysis. Finally, even though arbitrating between the differences of the stakeholders is a serious and sustained challenge, it is still not enough for properly governing risk. For example, all of them could all be objectively wrong about how key underlying processes work. What we would like to do is to properly govern risk, such that no matter the circumstances, all parties can attest that a good faith effort was made in managing all relevant factors.

Scenarios and Biases

Foresight has a standard approach for tackling the multiple counterfactual forests of different stakeholders, namely to assemble them into distinct scenarios or storylines that reflect plausible sequences of events as some set of driving factors prove to have stronger consequences than others. This scenario approach [Schwartz, 1991], most famously applied by Shell ([Team, 2005]), has come to develop sophisticated methods for eliciting and representing the kind of temporal multiplicity faced in these issues (in particular, see [List, 2004]). As such, this project looks at developing stakeholder-sensitive scenario construction, simulation, and usage practices suitable for the risk governance of high-hazard, long-term material commitments. Therefore, these methods are designed to discover both the causal structures underlying risky phenomena, both in terms of their objective effects as well as their subjectively perceived harms.

It is appropriate to question if the methods currently used within foresight provide the right cognitive toolkit for considering questions of risk [Verdoux, 2010]. Unfortunately, the answer appears to be negative, as scenario-based approaches can cause support-theoretic biases [Tetlock, 2005a], or the bias of assessing that a given event is more likely the more variations or reasons for it, no matter the likelihood of those variations or the quality of those reasons. In order to address this barrier, we turned to new research in causality developed in developmental cognitive psychology, which investigates how people form the structure of dependencies leading to their expectations. Using this framework, we can inquire not only into the predictions of stakeholders, but into the structure of dependencies between events perceived by the stakeholders, while not introducing additional criteria favoring particular options.

This Project: Its Purposes, Objectives, and Scope

Foresight, like history, requires one to structure information in a way that allows the reader to understand its narrative threads, but unlike history, the various strands that form its rich tapestry have not yet come together. To govern risk means discovering and keeping track of the tangle of mutually-inconsistent counterfactuals, and continually interrogating them for their consequences. This project has developed an approach designed for discovering the relevant facts and concerns of distributed risks and persisting them for continual use, in an online[1] fashion. This approach has three purposes:

- **Discovery** We would like for individuals attempting to govern distributed risk to identify and understand as many of the relevant factors at play as possible. Therefore, we have designed a way to elicit knowledge that assures that what we learn is usefully structured, but only includes the very weakest of preconceptions.
- **Knowledge Critique** We have no prior preconception of how rich or complete the mental models are, either with respect to empirical knowledge or with mutual understanding. By eliciting structure, we can discover where those structures are incomplete, inconsistent, impoverished, or vague.
- **Analysis** In the cases where stakeholders do have cohesive pictures of the scenarios facing them, it is entirely possible that the consequences of this understanding, when combined with the understanding of fellow stakeholders, is unclear. It might simply be impossible to understand the tangled web of consequences without aide. By formalizing and simulating these mental models, we may be able to build some intuitions about the aggregate effects of their combined scenarios.

Given those purposes, this project has the following objectives:

[1] By online, we mean online in the sense of 'online planning', or undertaken while the processes under study are ongoing, as opposed to meaning 'accessible via the internet'.

- Engineer scenario representation methods that allow for the capture, analysis, storage, and reuse of causal and impact information.
- Develop elicitation methods that progressively delimit and arbitrate governance deficits.
- Implement simulation methods capable of demonstrating plausible scenarios from elicited causal structures.
- Position uncertainty discovery as a valid governance need.

As important as it is to specify what it is that a project will do, it is equally important to say what it will not accomplish, and give a sense of the overall scope. While it would be ideal if this project analyzed a distributed risk infrastructure problem in depth, that would be a more appropriate scope for a doctorate. Instead, this project develops theoretical foundations for representation, elicitation, and analysis.

It also would have been better if the elicitation method presented in this report was subjected to field trials, but that was not undertaken due to circumstantial reasons. As it stands, the elicitation procedure should be considered untested, offered as a demonstration of a method that could meet theoretical constraints. It is also important to say that this project did not have a substantial public communications component nor an immediate commercialization of its findings. This choice reflects the personal temperament of the author, as either would have been suitable activities.

These methods of developed in this project are not above evaluation, and criteria for evaluation is given. Yet, the scope of this project is such that these approaches were developed theoretically, and thus should be considered unevaluated. As such, although you may find applications, there is yet no warranty, either expressly stated nor implied. We can say that given the preliminary nature of this attempt, it is extremely unlikely that the methods found here are optimal by those measures.

Methodological Framework

Overall, the methodological framework developed here works according to a layered approach (see Figure 1). In the innermost layer we find the exact distributed risk issues that we are trying to deal with. If these problems are not entirely unique, then we can learn from our mistakes and develop theories of what good governance and regulation consists of, and how it can fail. We then have a more general body of knowledge for governing risk issues in general, (next-to-innermost layer). As we will see later, these mistakes often consist of not looking widely enough, to discover all of the relevant phenomena, stakeholders, and options available. Therefore, one way to address these challenges is to use methods that are designed to be grounded in the insights of the participants (outermost layer). However, these methods are open to the extent that it can be tricky to establish that they address the discovery challenges posed in the risk domain. For that reason, we take on the insights of methods that pose additional domain-general inductive constraints (next-to-outermost layer), which we show here provide a structure that allows us to have some confidence that the design methods in question are addressing the discovery problems in question. In short, open methods that take on additional causal constraints can address the risk governance deficits often encountered in distributed risk issues.

Let us now look to the sources of this methodological framework. First, we describe how we have employed design research methods. Next, we describe why technical modeling is an appropriate stage for a discovery process. After this, we describe the technical approaches used here. Finally, we talk about how the discovery capabilities of design offer a new horizon for technical analysis.

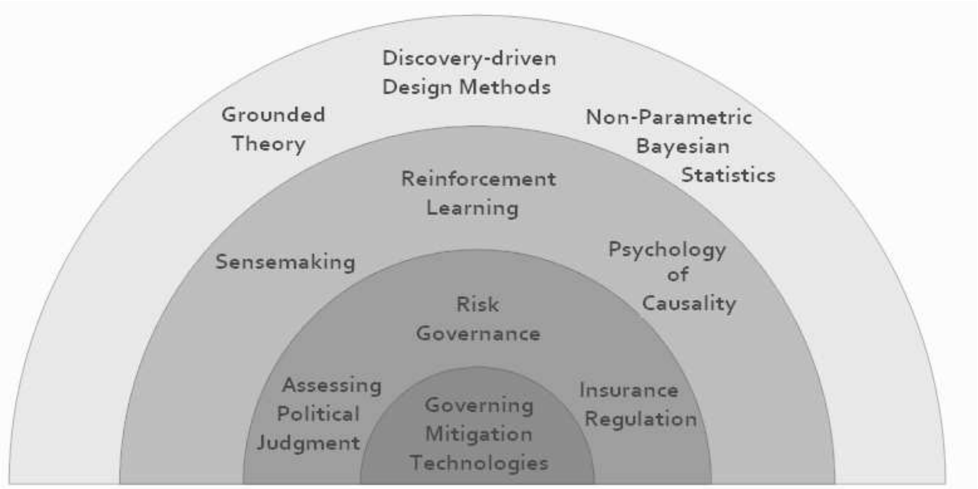


Figure 1: A Methodology of Layered Inductive Constraints

Design Research as Research for Discovery

We hope to provide scenarios solidly grounded in the input of the participants[2], at the risk of not initially making inferences we could otherwise posit if we were to assume background knowledge.

There are many ways to go about eliciting information from participants, but this work uses an interview protocol. Although interviewing is not always the most appropriate or engaging means of elicitation, it often entails a smaller time commitment for the participants, and is ideal for projects of tightly-controlled duration. Furthermore, by interviewing we can make sure that our participants does not collude, and thus assure that our results could have been gathered in any order, allowing us to make some assumptions in the mathematical analysis of our results. In any case, the criteria found through crafting interview protocols can then be applied to other elicitation methods.

As mentioned before, one of our central challenges is to sort out the objective and subjective perceptions of the stakeholders and observers. What we want to tease out

[2] We explicitly undertake neither the methodological approach of grounded theory [Glaser and Strauss, 1967] [Strauss and Corbin, 2008], nor explicitly adopt whatever ideologies lie at the root of the theory. However, there is something in the general spirit of the idea that's appropriate to invoke here.

specifically is how the stakeholders understand the benefits and risks present within the overall subject, as well as the risks and benefits of various activities within that subject. However, one key challenge is that different stakeholders might have vastly different cares and concerns. We need to discover what this mental model is, while at the same time directing their attention to key questions. Our protocols therefore aim at constructing their understanding of the causal structure of ongoing and future events, and their stakes within those events.

We only have a first-order approximation of how our stakeholders understand the issue at hand. Therefore, it is appropriate construct the elicitation procedure so that it will be as open-ended as possible. The open-ended interview [Fontanella et al., 2006] [Jarratt, 1996] [Kuniavsky, 2002] attempts to use only the words of the source, touching only on present experience. Unfortunately, given our problem, we can follow neither follow the path of inquiry developed by the interviewee with complete fidelity, nor can we focus on the present experience entirely, as we are trying to understand their perspective on future problems. In order to develop a grounded understanding of their future actions and concerns, we need to have the participant connect the current conditions to possible future conditions.

The observation that follows is that although a fully open-ended elicitation may not be appropriate, we can work in terms of broad, yet constrained, units of analysis. The Dervin sense-making interview [Dervin, 2001] offered us a demonstration of how to make generic questions that elicited the concerns of the stakeholders, as well as several key generic features of problems (obstacles, goals, bridges, actions) that could be carried into our own protocol[3].

To get around the problem of needing both understanding and coverage, the elicitation methods used for this project are iteratively deepening. In other words, we start with as high-level of an overview question as we can, and then proceed to ask open-ended model

[3] However, we cannot claim any similarity with Dervin's methods beyond these superficial features. In particular, this method is not necessarily critical.

eliciting questions (a broad sampling of examples is provided in the protocols section) until either the different aspects of the model (current behavior, eventual consequences) are joined or the participant admits that these are yet unconnected, at which point a more specific question is asked. Then, the goal is to build a connected model around both this question, and to the mental models outlined before.

Deepening can also occur across sessions, a feature this approach shares with the Delphi process [Linstone and Turoff, 1975]. There are numerous instances in which this may be appropriate to undertake another elicitation session. The reasons to do so might be for many different methodological objectives, from simply prompting stakeholders to think further to confronting stakeholders with divergent opinions. However, extreme care should be taken not to introduce these factors into the first of the interview, as to do so would compromise the criteria of avoiding support-theoretic biases, as well as violate exchangability, a statistical property necessary to the evaluation of overall elicitation completeness.

In our own model, we became concerned with the connectivity of actions: do current actions have any understood correspondence with the future risks? However, we found that we could ask open situation-posing questions in terms of participant-provided entities, that made use of an objective/subjective lens [Pennefather and Jones, 2008] [Pennefather and Jones, 2009] to build scenarios ready for analysis.

Why Technical Modeling?

> *"Welcome to the only game in town"* -Achewood [Onstad, 2009]

As this work focuses on building stakeholder-sensitive, causal models of risk and using them for simulations, it is worthwhile understanding what models are and why we would resort to building models. Although Miller and Page's definition of models [Miller and Page, 2007a] and their justifications for computational modeling [Miller and Page, 2007b] are valid and could serve us well, they do not address a design audience directly. Similarly, Bruce Bueno de Mesquita's justifications for applying mathematical models to political questions are salient for risk governance [de Mesquita, 2004], but yet still does not frame modeling practices in a way clearly acceptable to a trans-disciplinary audience. However, I have found a methodological framing that does suit both the character of this work and the audience, namely the use of structural formalisms as a clarifying device [Tilly, 2004]. Below I summarize this framing in my own words.

Models are formalisms for selecting particular aspects of the overall problem to be analyzed and conducting that analysis, which is merely seeing how those aspects combine or behave together under the assumptions that formalism entails. Model construction is a design activity in which that selection is undertaken, and where those selected factors are integrated. It is inappropriate to say that the results of a modeling activity should be taken as any final word, but they provide new insights that either suggest or eliminate particular design directions. We engage in modeling for the same reason we engage in all design research activity: to view those aspects of the problem in a way we could not otherwise.

Computational modeling has two singular benefits. The first is that it forces one to attempt to state the problem in a way that a computational formalism can use. This alone can uncover many aspects of the problem that were not thought through to the level the formalism may require[4]. The second is that the resulting model will frequently behave in an unexpected way, that then requires an explanation, and therefore uncovering new insights about the problem[5].

However, there is an even more fraught observation that must be made about modeling: it is, to follow the Achewood quote, the only game in town. We are all limited to

[4] Having said that, it is not infrequent that one discovers that a particular method of computational modeling is unsuitable for the problem at hand. No method is suitable for all stages of all problems.

[5] Again, it is also the case that problems in the result of the model can also indicate problems with the suitability of the formalism employed. It is a risk governance deficit to put too much, or too little, trust in models.

where we can reach given our personal limits of observation and cognition. We have to accept that, as inclusive as we may try to be, we will never take in the whole of another's experience. We are left to operate on what we know, and therefore having many modalities for discovering our own limitations and inconsistencies can give us a better sense of what is appropriate. Viewing the frailties of our assumptions is a way to gain some humility and perspective on our own understanding.

Technical Methods

Given this framing, we can now introduce the technical methods we are using in proper context. As a starting point, we turned to the paradigm of reinforcement learning [Sutton and Barto, 1998], also called neuro-dynamic programming [Bertsekas and Tsitsiklis, 1996], and is widely used in applications such as robotic-motion planning [LaValle, 2006]. These methods were selected for their general, yet directly applicable, units of analysis: states of affairs, actions, stakeholders, observations, rewards/losses of varying criteria, and discount rates. At times, the similarity between the abstract domain of these methods and those of risk governance is uncanny. For example, using the method of *value iteration*, one could say rather simply that closed-loop supply chains can have an unboundedly higher reward that those that consign their materials to the scrapheap at the end-of-life.

These models are nearly appropriate, but need to be augmented in various ways. The first is that the preferences of each stakeholder are dependent upon the current state-of-affairs, but very often there is no distinction made between the external state-of-affairs and the state of the stakeholder themselves. Therefore, we wish to be very explicit in specifying that the agent is not proceeding according to some static set of utility evaluations, but instead may change their preferences, but are not dictated to do so, according to a variety of factors including sociological pressures. Given the pluralistic nature of risk governance, just as we do not reduce every perceived risk to objectively verifiable harms but will welcome these simplifications when they are available, we do not demand that every such pressure be decomposable into psychological effects but will certainly welcome those decompositions whenever such conclusions are found within analytical sociology (such as in [Hedstrom, 2005]) or can be attributed to structurational effects [Orlikowski, 1992].

The abstract, control-theoretic states-of-affairs provided by reinforcement learning actually already have a powerful connection to the representation of time, due to their foundations in Markov chains [Norris, 1997]. Some notions that they borrow include memory, rates of change, irreversibility, mixing, and entropy. The actual implementation of the simulation is done with conventional approaches [Law and Kelton, 2000].

Design as a Discovery Process

Early on in the SFIN program, I became suitably impressed with the ability of design methods to discover useful insights from qualitatively rich information. Although some of my colleagues trained in more classical statistics questioned the validity of this work, I was rather more impressed. This is because, in particular cases, the observations generated were not just data, but all of the underlying conditions that were observable about why that phenomenon was the case, and not some other. There existed a distribution where this myriad of factors was able to come together in a way that was explainable, and where few other counterexamples presented themselves. For an illustrative example, one could do worse than to consider Go-Gurt [Squires, 2002].

Go-Gurt was created with the aide an anthropologist who stayed with a two-child family as that family was waking up early to go to school. One of the children in the story, despite the mother's concern to serve a healthy breakfast, is not hungry at breakfast time. This lack of hunger, given the physiology of small children, the hour of the day, and the short period of time after waking before breakfast, is not surprising. This child would then eat their lunch earlier in the day, leaving himself hungry in the afternoon.

Given this single observation, we should not be surprised that a nutritious, fun-to-eat, portable, breakfast product shaped in the familiar form of a banana shape served that family's needs well. I can tell you that both early schooldays and child physiology are far from peculiar to this particular family. What this implies is that the mere fact that this happened once, combined with prior knowledge about life in the consumer market generally, should (and gladly did) make the commercial prospects of a market intervention strongly salient. A Bayesian statistics capable of transferring priors and capturing the causal salience of single insights would be a powerful tool indeed for ethnography and other design research approaches.

This view is also shared by a recent article about design education by Don Norman [Norman, 2010]. One does not have to agree with the polemical thrust of the article's application to either design or design education, found within there was an interesting proposal to statisticians: investigate statistical methods suitable for the way designers work.

> *Designers are practitioners, which means they are not trying to extend the knowledge base of science but instead, to apply the knowledge. The designer's goal is to have large, important impact. Scientists are interested in truth, often in the distinction between the predictions of two differing theories. The differences they look for are quite small: often statistically significant but in terms of applied impact, quite unimportant. Experiments that carefully control for numerous possible biases and that use large numbers of experimental observers are inappropriate for designers.*
>
> *The designer needs results immediately, in hours or at possibly a few days. Quite often tests of 5 to 10 people are quite sufficient. Yes, attention must be paid to the possible biases (such as experimenter biases and the impact of order of presentation of tests), but if one is looking for large effect, it should be possible to do tests that are simpler and faster than are used by the scientific community will suffice. Designs do not have to be optimal or perfect: results that are not quite optimum or less than perfect are often completely satisfactory for everyday usage. No everyday product is perfect, nor need they be. We need experimental techniques that recognize these pragmatic, applied goals.*
>
> *Design needs to develop its own experimental methods. They should be simple and quick, looking for large phenomena and conditions that are "good enough." But they must still be sensitive to statistical variability and ex-*

> *perimental biases. These methods do not exist: we need some sympathetic statisticians to work with designers to develop these new, appropriate methods.*
> From *Why Design Education Must Change?* by Donald Norman

Although this project only directly addresses scenario formation processes in the domain of risk governance, I have hopes that analytical techniques for causal discovery might also apply to foresight, design, and ethnographic activities.

Background Research

This project hopes to draw on the wisdom and insight of many different fields. As such, the literature review for this project engaged the coarse material of survey papers, textbooks, and established programs of research as resources, engaging current questions only as necessary. We have already been introduced to the framework of risk governance, which serves as both a context for the overall study, and through risk governance deficits as guidelines that define the objectives of the methods discussed here. Similarly, we have already introduced design research methods focused on elicitation and sensemaking and technical methods of modeling multi-stakeholder decision processes under uncertainty, which provide us our methodological and analytical backgrounds respectively.

With this backdrop, we can now focus on the core problems that this research engages (see Figure 2). This research addresses the challenges posed to foresight by key set of studies in the statistical quality of political expertise, which found terrible results in terms of classical Bayesian norms ([Tetlock, 2005a]). This research occurs at a time in which venerable approaches for assessing psychological performance to classical norms [Tversky and Kahneman, 1974] is being brought into question using a new framework of causal discovery [Krynski and Tenenbaum, 2007]. This work seriously engages the possibility of using causal Bayesian norms for issues of political judgment.

Following this section, we will introduce knowledge vital to understanding using causal

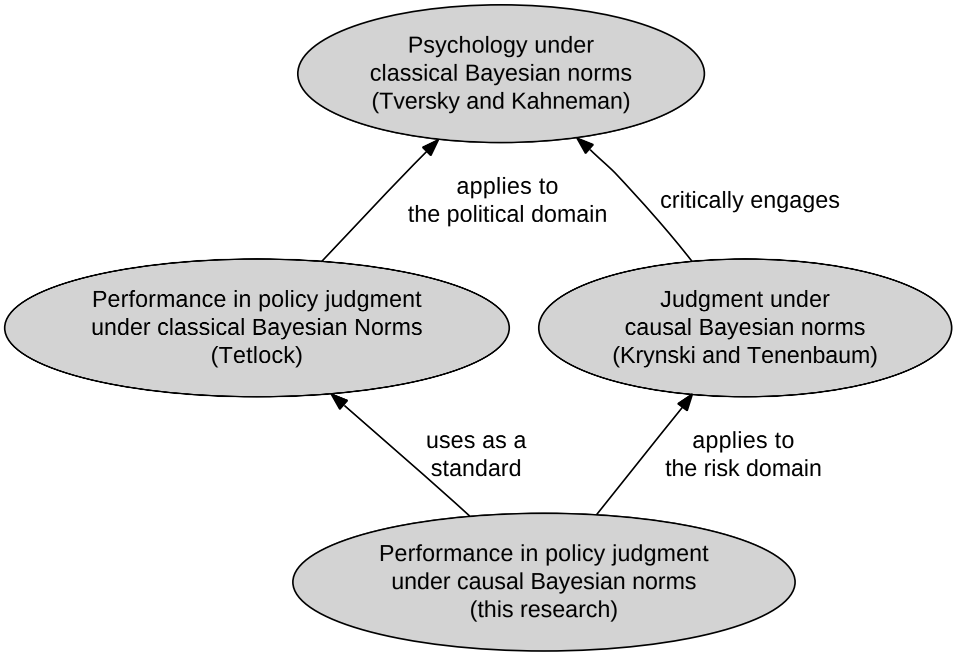


Figure 2: Core Research Engagements

approaches in the risk domain, including basic introductions to causality and catastrophic risk policy. Finally, we will look at the policy challenges posed by distributed risk issues specifically.

Statistical Model Elicitation

> *"What is it about politics that makes us so dumb?"*-Philip Tetlock, quoting Daniel Kahneman [Tetlock, 2007]

Although there have been attempts to elicit statistical models in complex areas such as political science [Gill and Walker, 2005] and ecology [Kuhnert et al., 2010], a very ambitious and substantial research program has cast doubt on the applying these practices to policy. In *Expert Political Judgment* by Philip Tetlock [Tetlock, 2005a], he shows that by certain statistical norms, political experts rarely show a predictive understanding of their areas of concern better than predicting the most recent rate of change, that the experts who are most consulted by those in power and by the media predict as badly as random choice. Further, the experts who are regularly consulted despite doing the worst are those that are most confident, are the most reluctant to change their further estimates when some are shown wrong, are the least likely to entertain alternative possibilities, and are so set in their views that even good historical evidence would be unlikely to change their minds. Perhaps our more cynical side would not be impressed by the possibility that qualified nuances do not engage our emotional registers in the same way as indignant certainty, but it is unfortunate to have that sad intuition so broadly confirmed.

One striking factor about these findings is that systems thinking is the root of both good and bad predictive judgment. For good predictive judgment, it allows for the possibility of forces to come together in a variety of ways, as various trends come to dominate or be squelched by factors that occasionally have very tiny thresholds, and have a good counterfactual imagination. Those with bad predictive judgment see an unequivocal combination of reinforcing and balancing factors which will drive a particular contingency while mitigating others. Therefore, the capable expert not only has to think systematically, but must also be able to integrate multiple perspectives effectively.

Perhaps the most troubling result for foresight practitioners is that scenario exercises, which are explicitly designed to boost the counterfactual imagination, fail to change the mind of those with set beliefs, and distort the opinions of the open minded; in short, making the best worst off while offering no improvement to the rest. This distortion happens through the introduction of support-theoretic biases, which increase the perception that a given factor is more likely given the number of reasons listed to support it. In the experiment described, experts were first to estimate the likelihood of violence breaking out throughout the Cuban missile crisis of 1962, and then were to consider various scenarios in which different kinds of violent conflict broke out. The result left the opinions of those who resisted counterfactuals unchanged, while it left the results of those willing to consider counterfactuals logically inconsistent, giving probability estimates of the various events that summed to higher than one.

Admittedly, scenario exercises can be constructed in a variety of ways, and introducing support for scenarios that were previously unconsidered can reduce the support for the status quo, perhaps yielding yet broader experts. Nonetheless, it would be troubling if the results of scenario exercises could be coolly manipulated by merely controlling the number of facts considered. I can appreciate some views that would indicate that we should not be concerned; in particular, it could be tempting to say that prediction is an invalid measure of anything, and that the kinds of stories that develop visions produce change-makers who will have a normative impact despite the content of their objective beliefs. However, while that may be appropriate for some design activities, the mitigation of risk governance deficits does involve being able to realistically integrate and give simultaneous consideration to multiple conflicting accounts of objective and subjective risks.

However, before once again riding this train to the inevitable destination of the objective/subjective divide, let us consider something that should give us pause: are the findings using the appropriate objective standard? I have always been concerned about the framing of biases: according to exactly which model are particular inclinations biases to? When are biases actually artifacts of using different statistical models, which may be more appropriate for navigating the risk environment actually faced by a given stakeholder? This may be an especially appropriate question for political decision making, where many core events are remote possibilities and that the appropriate handling of which requires making sound trade-offs from limited information. In order to assess if this is correct, Tetlock gives us a demanding but fair standard:

> *Promoters of "debiasing" schemes should shoulder a heavy burden of proof. Would-be buyers should insist that schemes that purportedly improve "how they think" be grounded in solid assumptions about (a) the workings of the human mind and -in particular- how people go about translating vague hunches about causality into the precise probabilistic claims measured here;*

> *(b) the workings of the external environment and -in particular- the likely impact of proposed correctives on the mistakes that people most commonly make in coping with frequently recurring challenges.*
> From *Expert Political Judgment* by Philip Tetlock [Tetlock, 2005b]

It is the hypothesis of this work that newly developed causal Bayesian statistical frameworks offer a better explanation of human decision making, while also being more suited for many kinds of realistic problem solving activities than "unstructured" statistical techniques, which would not be able to converge to any sound assessment based upon the limited sample sizes found in may problems posing risks [Krynski and Tenenbaum, 2007]. Indeed, it is to some extent a testimony to the existing body of expert knowledge that the right variables are available to be provided to the statistical methods, enabling their steady performance.

However, to be clear, it is not the position of this work that bad statistical performance should readily be excused as a result. I would prefer to think that those who were the most nuanced, most skeptical, and already had the best performance would be in the best position to incorporate the statistical models that 'beat' them into their own toolkit, and would use them in an appropriate and integratively sophisticated way. Further, Tetlock's work has done the most admirable job in developing frameworks to keep experts accountable to reality, and any criticism here should be taken as an attempt to extend this work, instead of undo any of the foundations it offers.

As an example of one such extension, it would be useful to be able to not only consider cognitive habits of those who show good political judgment, but to also discover the characteristics of good cognitively-realistic models on their own terms. Is it true that good expertise is formed by deferring judgment between mentally maintaining the competing equilibriums between distinct counterfactual paths, while maintaining an ultimate skepticism that gives boring but reliable trends their due?

Learning Causality and Discovering Structure

> *"If we're so dumb, how come we're so smart?"*-Clark Glymour [Glymour, 2001]

How is it that the activities done in design might function in reducing risk? One answer is that design processes may discover previously unknown stakeholders, criteria, perceptions, and influences in risky processes. Surely the processes most associated with the creative (and therefore valuable) aspects of design, such as brainstorming and prototyping, provide exactly this capacity for discovery. Other design processes, such as ethnography, usability studies, and observation, provide a small number of samples with a very rich feature set, demonstrating that certain factors dominate others in the way that systems interact, and that at least in certain scenarios, we can say deterministically that a certain path of cause and effect, as well as certain interpretations of them, was observed.

Given this, what we want is a design statistics that effectively learns gross distinctions from a small number of observations, that successfully incorporate an unknown number of previously unknown distinctions and criteria, that assembles 'theories' of interaction that transfer as correct biases for related situations, and builds theories of the results of interventions that are accurate enough to justify considerable expense in testing them. It is remarkable that these are exactly the problems that must be solved by infants and children in learning how to classify objects, learn languages, understand the consequences of actions, and act socially. Perhaps it is not entirely surprising, given that both the beginning of a project and the beginning of a life are both events that necessitate rapid learning. In any case, even if there does not prove to be any actual connection in the kinds of learning taken at these respective early stages, that is no reason not to borrow methods designed to address these questions. Therefore, methods from statistics that have proved useful for cognitive developmental psychology may also be useful for design applications.

First of all, let us consider how to build models that successfully incorporate an unknown number of previously unknown distinctions and criteria. This is very important to risk governance, as we do not know whom all of the stakeholders are, or how much the concerns and criteria of various stakeholders overlap, all of the different causal factors at work, or in fact the appropriate extent of the scope of the problem, generally speaking. However, there are now statistical processes which can flexibly accommodate learning an unknown, and potentially infinite, number of parameters, namely the field of non-parametric Bayesian statistics (as nicely surveyed in [Jordan, 2010]). The models from this field can handle unknown numbers of categories [Aldous, 1985], infinite numbers of potential overlapping features [Griffiths and Ghahramani, 2005] [Thibaux and Jordan, 2007], and even distinguish previously unknown hierarchies of shared features in potentially infinite pools of features [Miller et al., 2008]. Let us now see how this statistical work will help us.

In the case of this study, we might always yet discover new concerns, new stakeholders, and new events affecting our existing assessments. Yet, at the same time, we have reason to believe that whatever work we have already done is effective at discovering what we have asked. As we ask more stakeholders, we expect that we will first discover the most widely known phenomena and most widely shared interests, and as we continue to ask we will continue to hit pockets of less known but still significant phenomena and interests, but will progressively find diminishing returns as the number of sources elicited from grows. We could say that there are possibly an infinite number of stakes held but they become increasingly tangential, and at some point we have no practical way of knowing what we are yet excluding.

An intuitive model for this distribution of discovery is the Chinese Restaurant Process (CRP), in which we suspect that each concern discovered will either be a previous concern, chosen randomly but distributed in proportion to the concerns discovered so far, or occasionally a new category, found decreasingly in proportion to the number of samples already taken, with the ratio of new elements in initial proportion to the term γ. The CRP is named after a genre of restaurant consisting of a (seemly) infinite number of large tables, in which new restaurant goers join tables in proportion to the number of people already sitting at them, as those tables are likely to have the best food or party members known to the arriving guest, but will occasionally strike out and head to a new table. Here is a mathematical expression for the CRP (a guide to the mathematical notation used in this paper can be found in Appendix D):

$$P(z_i = k|z_1, \ldots, z_{i-1}) = \begin{cases} \frac{n_k}{i-1+\gamma} & \text{if } n_k > 0 \\ \frac{\gamma}{i-1+\gamma} & k \text{ is a new class} \end{cases}$$

It turns out that the random process underlying the CRP is the Dirichlet process, $\mathcal{DP}(\gamma, G)$, where G is a permutable measure. What this means for us in practical terms is that as long as our sampling process is exchangeable (or, in other words, as long as we do not have any ordering dependencies in which we sample), we can still trust the statistical validity of our results without making any assumptions that the samples are identically distributed. Further, each of the underlying categories can define distributions of their own, leading to a Dirichlet process mixture model, which is to say, a set of statistical models who's memberships are distributed according to the CRP. We will say a category z is distributed CRP with an exploration γ through the following notation: $z|\gamma \sim CRP(\gamma)$.

What if instead of expecting to slowly discover categories, with each new sample we expect to discover some factors. Fortunately, that too can be accommodated by the Indian Buffet Process (IBP). In this process, we pick some number of factors in proportion to those previous samples that already have it, selecting some number of new factors otherwise. It is as though one were going from seemingly infinite buffet of Indian food, trying dishes as one sees earlier guests trying them. We can say that, we expect a distribution Z of the n^{th} sample over k factors to be distributed as a Bernoulli trial in proportion to the factors found in previous samples along with a Poisson over the number of new factors[6], which is expressed mathematically as follows:

$$Z_{n,k}|Z_{\overleftarrow{n},k} \sim \begin{cases} Bernoulli(\frac{m_k}{n}) & \text{where } m_k > 0 \\ Poisson(\frac{\gamma}{n}) & \text{number of additional new factors, } k \end{cases}$$

It has turned out that this IBP also has an underlying process that permits exchangeability, namely the Beta process. Like the Dirichlet process, the Beta process can also serve as the basis of a mixture model.

One important feature of mixture models is that they can be assembled hierarchically through trees of conditional inference. For example, given that we have identified a given kind of stakeholder, say an insurance company, then we know it is likely that they have some concerns and not others, though this is only a prior since insurance companies are not homogeneous. We can also have a prior over the spread within a given kind of stakeholder: for example, insurance companies are probably more homogeneous in their concerns than protesters. In short, this implies another of our attributes for a good design statistics: it should learn 'theories' of interaction that transfer correct biases for related situations. Hierarchical non-parametric Bayesian models have been demonstrated to learn this kind of theoretical homogeneity [Kemp et al., 2007], and have been extended to discover complicated qualitative structure, including medical ontologies, aboriginal Australian kinship systems, and political alliances [Kemp et al., 2006].

When combined with a very tiny amount of prior capability, such as the ability to form graphs, grammars, or predicates, the capability to transfer knowledge implies the capability to effectively learn gross distinctions from a small number of observations, which was another of our desired features of a design statistics. Simple graphical constraints make it possible to learn complicated relationships, such as clusters, spaces,

[6] $\overleftarrow{n}$ should be read as "all samples up to, but not including, the current sample."

grids, chains, grids, and rings; for example, that animals form a tree-like hierarchy and that the distances between world cities can be well described by a grid of rings [Kemp and Tenenbaum, 2008].

Perhaps most relevantly for this work, a statistics suitable for foresight should account for the ability to learn reliable theories about the result of interventions, even if these interventions have not been taken, and remain 'counterfactual'. Developmental psychologists have posed the possibility that children, at a very young age, learn mental representations called 'causal maps' [Gopnik et al., 2004], which allow them to make sensible inferences about the results of possible actions, without actually having to undertake them. These results follow a near revolution in the philosophy of causality and its applications in statistics and computer science [Pearl, 2000] [Spirtes et al., 2000]. This work finally puts common intuitions about the difference between causality and correlation into a workable framework. For example, we should be able to capture such clear intuitions that tampering with your barometer does not mean you can control the weather, nor does accidentally setting off your burglar alarm imply someone is robbing your house.

Beyond the success this work has enjoyed serving as an experimental research paradigm for developmental psychology [Gopnik and Tenenbaum, 2007], this has important consequence for risk governance: even in the policy sphere, we should expect that while some interventions cannot be easily untangled from complex webs of mutual interaction, while others clearly have no substantive dependence on earlier events, while having strong downstream consequences. Consider the following examples of counterfactuals from *Expert Political Judgment*:

> *"If the carriage driver of Archduke Ferdinand had not taken a fateful wrong turn that gave the Serbian assassins a remarkable second chance to carry out their previously botched assassination plot, war would not have broken out in August 1914."*

> *"If Stalin had lived several years longer (surviving his stroke but in an irrational state of mind that encouraged high-risk adventures), World War III could easily have broken out in the mid-1950s."*
>
> *"If bad weather had delayed the discovery by U-2 reconnaissance planes of Soviet missiles in Cuba until the missiles were operational, the Soviets would have refused the American demands to dismantle and withdraw the weapons."*

It seems entirely appropriate that each of the antecedent in these statements should be judged independently of how politically likely their consequents are assessed. In other words, despite how likely different political ideologies assess these statements overall, we should find no correlation between political preference and the assessed likelihood of a lost driver, an individual medical outcome, or the vagaries of the weather, and the fact that Tetlock finds such correlations is strong testimony to the problems of bias in political expertise.

The consequence of all of these factors taken together is that, given domain-general inductive constraints, causal effects can be learned effectively [Goodman et al., 2008b] [Tenenbaum and Griffiths, 2003] [Griffiths and Tenenbaum, 2007] [Kemp et al., 2010]. Therefore, given the loose units of analysis defined by the overlap of risk governance and reinforcement learning as inductive categories, the challenge is to develop elicitation strategies that effectively learn the causal models of the stakeholders, allowing for a clear understanding about the perception of interventions, and thus lead to the mitigation of risk governance deficits.

A Very Brief Introduction to Causality

Given the fundamental role causality has in this research, it is important to review the basic concepts and definitions [7]. Causality, for our purposes, is the study of the dependences between events, and how interventions change those dependences. If I were to ask you "What is the likelihood of it raining on any given day in Toronto next year?", the answer would be something along the lines of the percentage of days it typically ranges in a given year, perhaps weighted more strongly by recent years to account for changes in the local climate. On the other hand, if I were to ask "What is the likelihood of it raining in Toronto on a spring day next year?", a better answer would likely no longer be the yearly average, but an average taken over days in spring, as we know that the weather depends upon the seasons. For this reason, we might say that the first is $P(\text{rain})$, and the second is $P(\text{rain}|\text{season})$, which we take to mean "the probability of rain" and "the probability of rain, depending upon the season" (although both are implicitly given the ground conditions of being evaluated per day in Toronto). Even in the case of the first question, I might might be able to provide a better estimate through knowledge about the proportion of the year that each season lasts, and the likelihood of rain in each season, by adding up their respectively likelihoods in proportion, i.e. $P(\text{rain}) = \sum_{season}^{Seasons} P(\text{rain}|\text{season})P(\text{season})$.

Sometimes phenomena have no detectable paths of dependence between them, such as whether it is raining ($P(\text{rain})$) and if I receive an email from my friend in Japan $P(\text{email from Japan})$ on any given day [8]. Then, we would see that the probability of rain is not influenced by the email, ($P(\text{rain}) = P(\text{rain}|\text{email from Japan})$), and we could say that these two events are independent ($\text{rain} \perp\!\!\!\perp \text{email from Japan}$).

Although the likelihood of one event is useful information about the likelihood of another

[7] See [Pearl, 2000] for a thorough introduction

[8] In some ways, this example already demonstrates the inescapable nature of background conditions, such as to divide units of time in comparable ways, or to enjoy the perceptual richness to distinguish both of these phenomena. We will, as throughout much of the rest of this paper, leave these questions to philosophers and psychologists, and instead enjoy the fruits of their work uncritically.

event, it does not mean that this event causes or is caused by that event, as these two events may have some common cause. Consider that classic Midwestern observation, "Every time you wash your car, it sure seems it's more likely to rain." In this case, these events have a common cause, namely the time since it last rained, which was long enough for the car to get sufficiently dirty as to have needed it. So, these two events, while often having an uncanny correlation, do not cause each other, and are independent given their common cause (rain $\perp\!\!\!\perp$ wash car|days since last rain).

Now suppose that I go out on the back porch for breakfast, and find that the garden has been watered, or P(garden watered). With negligible exceptions, I know that this is a result of either that it had rained (P(garden watered|rain)), or that my wife ran the sprinkler after I went to bed (P(garden watered|sprinkler)). I may be interested to know whether it rained (P(rain|garden watered)) or whether the sprinkler was run (sprinkler|garden watered).

Now suppose I learn from my wife that she had run the sprinkler. As a result, the watered garden has been explained, and I now presume that it probably did not rain, or at the very least that the likelihood that it rained overnight is no greater than usual. Although the likelihood that it rain given that the garden has been watered does not change substantially (P(rain|garden watered)), it certainly changes in this case! The use of the sprinkler is then said to "explain away" the fact that the garden has been watered. In general, when an intervention is taken, we will know that this particular cause offering its full contribution to the immediate effects, and the knowledge that it is been undertaken will reduce the knowledge those effects might give use about the likelihood of any other potential causes.

When we intervene, we 'sever' the path of dependences between events, such that other possible causes of the event are as likely as they were if the factor we intervened upon had no relation. For example, in a town in central Illinois, they test the tornado sirens on the first Tuesday of every month at 10 am. Given a tornado siren at this time, a tornado is no more or less likely than it would be on any day at 10 am, and thus we can say that this testing severs the dependency between the tornado and the tornado siren.

Causality is study of how dependences change based upon interventions. Causal dependence is different than the statistical definition of dependence. I may notice with a strong probability that I am too hot when a thermometer is above a certain level, but it would be ineffective to cool down by tampering with the thermometer. Although correlation is not causation, causation can be discovered through interventions.

Systems can change drastically upon interventions. Although a system may counterbalance the effect of an intervention, it is equally true that an intervention may sever the flow of events leading to reinforcing and balancing events, changing the dynamics of a system entirely. If we understand the causes that might bring about an intervention regularly, it is appropriate to describe a larger system that includes it, to which statistical dependence then applies. Yet it is equally true that the number of potential system topologies is combinatorially large, and even among the systems that we have observed any number of causal topologies are statistically indistinguishable without intervention [Spirtes et al., 2000]. Further, the combination of multiple paths of intervention would undermine otherwise straightforward changes in the dynamics of a 'single' system.

Risk Policy Basics

> *"Really, the risk to each of us is very small. At worst, we lose our lives."*-
> William T. Vollman

If we are going to understand how to govern catastrophic risks, we need to understand the policy concerns specific to catastrophic risks. For this, I turned to *Catastrophe: Risk and Response* by Richard Posner [Posner, 2004] and *Worst-case Scenarios* by Cass Sunstein [Sunstein, 2007], which are fine guides to catastrophic risk policy issues written for a general audience. These volumes cover many of the challenges encountered when placing the ethically imponderable into the realm of the legally actionable, such as the effective monetary value of lives, appropriate discount rates for different contexts, and precaution versus estimation under extreme uncertainty. This section also refers to content from *Risk Governance* [Renn, 2008] and from Mark de Figueiredo's Ph.D. dissertation [de Figueiredo, 2007].

Criteria for Risk Regulation

Considering the imponderables where we had left them, it is clearly impossible to have an infinite degree of precaution, as risks are an inherent part of life. We cope with this by treating risks differently: some we deem normal, others tolerable given measures for protection and risk reduction, and yet others are intolerable. Yet, how do we decide when risks are appropriate, or in other words, which risks should be subject to regulatory measures?

One standard is the demonstrated willingness to pay to avoid risks. Willingness to pay may be determined by observing all of the conditions in which a population currently does pay to reduce their risk, or demand compensation for increasing their risk, and aggregate these situations to determine a cost for a given risk. These studies reveal complications which do not allow a strict price for mortal risk.

First of all, the price that people are willing to pay to avoid risk declines faster than the likelihood of the risk. If we take each risk on its own, then this is not very sensible, as this preference pays for larger risks disproportionately larger than smaller ones. However, I have had an independent idea on this subject. Suppose that the number the number of risks grows as we look down the scale of likelihood. This is a reasonable assumption, as there are progressively more paths between chains of remote events at lower degrees of likelihood. Further, suppose that we can only afford to eliminate some number of them. In this case, picking the optimal combination of what we can afford is equivalent to the 0-1 knapsack problem, and thus is computationally hard [Karp, 1972]. If evaluating these possibilities takes some expense, then it would be infeasible to expect people to pay proportionately, and instead pricing more likely risks disproportionately higher would be a reasonable heuristic for achieving a balanced overall protection. Given this complexity, determining the actuarial cost of risks should be recognized as a valuable activity, worth compensating a firm or department for undertaking on one's behalf.

Another consideration is a consequence of risk being undertaken socially. Consider two risks that pose equal risks on an individual basis, but in one case, the risk to individuals is dependent, while in another case, it is independent. Surely we would like to pay more to prevent the dependent case, as the losses imposed on society are greater.

Even taking into account social and heuristic considerations for risk perception, the actuarial conception of risk still does not capture many relevant factors. The first of all, it is often not normatively appropriate to assess different risks identically. First of all, there is a difference between risks that individuals voluntarily agree to, and have some responsibility in controlling, versus those that they are subjected to without their consent. When risks are imposed, people distinguish between whether it is imposed by other individuals, who are presumably profiting from it, or whether the risk is natural, and thus is to the advantage of nobody. People also distinguish between the kinds of harms that risks impose, whether they are physical, financial, or otherwise. One way to get a handle on the resulting conception of risk is to consider some semantic categories for risk, and consider how people respond to them differently:

- **Emerging Danger (fatal threat)** Industrial facilities and other installations can break down, imposing a random and catastrophic risk on the surrounding population. Such facilities may be run with profit to their owners, where it is uncertain if that profit will be used to compensate the surrounding population in this case of catastrophe. People will demand high reductions to the risks imposed by this situation, but will have a low willingness to pay for this reduction without some share of the benefit.

- **Stroke of Fate** Natural disasters are posed as risks beyond human control, quirks of fate or acts of God. Due to rarity, individuals perceive natural disasters as proceeding according to natural cycles or to divine purpose, instead of specific random processes. For this reason, people will underestimate these kinds of risk, and will, for example, return to areas that are prone to flooding. This perception has made it politically difficult to mitigate against natural risks, leading to regulatory choices that are difficult to justify in areas such as flood insurance [Holladay and Schwartz, 2010].

- **Personal Thrill** In these activities, people directly undertake activities knowing that they are risky, and that they will have to exercise skill to overcome these risks. In order to be considered appropriate, these activities are completely voluntary, involving only those who agree to (and usually pay others for) them.

- **Gamble** In gambles, people make undertake risks in order to gain potential rewards. Gambles for any stakes other than monetary gains and losses are seen as ethically troubled. Gambles are characterized by probabilistic thinking.

- **Indicator of insidious danger (slow killer)** These risks are related to unobservable dangers found in air, water, and food. These risks are effectively invisible, and often cause their harm slowly over a long period, such that there is no personal sensation relating to these harms. Individuals are often willing to accept a reasonable level of risk given that they trust the institutions responsible for monitoring and regulating them. However, if such trust is lost, then individuals will demand very low risks in these areas, and that this reduction in risk is independently verified.

Other semantic categories may exist. I suspect that there are a large class of risks related to the conception of everyday life, in which individuals are expected to take and personally manage risks through skill in order to be considered competent in society. Such risks might include driving, shopping for food and other essentials, and risks related to conventional professions and workplaces. In all cases the political question of who should pay to reduce a risk affects how much any individual is willing to pay. Under the expectation that risks may be fairly adjudicated, this may be a rational response to risk when considered in aggregate.

Finally, there are times in which risk preferences reflect neither actuarial nor normative conceptions. In many of these cases, it may genuinely be the case that destructive biases are at work. One such bias is the availability bias, in which events that the individual experienced or are familiar with are evaluated as much more likely than their naturally occurring rate. As an extreme example, consider that following the 9/11 terrorist attacks, airline travelers were willing to pay more for flight insurance against damages due to terrorism than flight insurance against all causes (including terrorism). These biases, outside of their role as heuristics for social costs and social norms, represent vicissitudes which effective risk governance institutions must be able to resist.

Distributed Risk Basics

Distributed risks are created by problems that appear to have scientifically-uncertain irreversible long-term public costs that span borders, while the mitigation to those problems appear to have uncertain rewards and heavy short-term individual costs under specific regulatory frameworks (for an easier comparison, see Table 3).

Table 3: Disparities for Distributed Risk Phenomena and Mitigation Approaches

Risk Aspect	**Phenomena**	**Mitigation**
Responsibility Bearer	Public	Private
Magnitude	Scientifically-uncertain	Untestable
Temporal Extent	Intergeneration (unlimited)	Generational
Regulatory Regime	International	National
Spatial Extent	Worldwide	Unknown
Liability	No Retroactive Basis	Potentially Retroactive
Scope of Impact	Total	Incomplete
Temporal Origin	Uncertain	Immediate

Let us briefly consider two distributed risk issues. The first of these is the adoption of closed-loop supply chains as a remediation measure for raw material depletion. In its most basic form, raw material depletion is simply a restatement of the third law of thermodynamics, in that it is easier to convert an ordered or pure substance into a disordered or impure substance than vice-versa. Although we may be far away from these physical limits, it is safe to say that it is unclear if viable technologies for the restoration of materials can be figured out well enough to keep supplying demand levels that are hard to move away from. Collectively, we may not be clever enough to solve these thermodynamic problems, in which case we might suffer from an "Ingenuity Gap" [Homer-Dixon, 1995].

One proposed remedy has been to develop systems for maintaining and upcycling technical and biological materials, creating a closed-loop that does not suffer the thermodynamic decline so readily [McDonough and Braungart, 2002]. Unfortunately, the usage side of the loop currently introduces uncertainty in the rate at which resources will be available, posing short-term logistical challenges to closed-loop practitioners in comparison to their peers using conventional supply chains [Pochampally et al., 2009]. Thus, we see costs imposed internally to firms who attempt mitigation, while the bulk of participants endure longer-term supply risks, which in turn depletes the common pool of raw resources, leading to mitigation attempts being negligible in their overall effect.

As raw resources become scarcer, that also means turning to resources that are more difficult to discover and extract. As discovery becomes more difficult, there is progressively larger uncertainty about how much of a resource exists at all, such that the error in the estimate is a greater percentage of the tail of resources available [Taleb, 2011]. The onset of material scarcity may be diffuse, masked by unknown reserve levels in politically closed countries. Having said that, I have not yet convinced myself that an economic treatment of scarce resources will not arbitrate an equilibrium, and Posner's dismissal of the issue [Posner, 2004] has given me pause.

Another interesting distributed risk issue is the effects of climate change versus the measures designed to address greenhouse gasses. Imagine that you own an airline. An increase in the incidence of severe weather could present any number of problems: a greater risk of weather-related dangers when flying, more volatility in flight times and routes, hail damage to aircraft, extreme temperature wear on supporting airports and runways, a larger number of extreme weather events such as tornadoes and hurricanes, and greater insurance premiums against all of these concerns.

Now, as an airline owner, let us think about a different problem. Aircraft currently have few technical choices when it comes to emissions. Concerns about these emissions include regulations affecting the level of allowed pollutants, public expectations about the current efforts to reduce them, the appropriate level of investment into alternative fuels is appropriate, offsetting activities and how they are offered as part of the business, future liabilities for current pollution activities, and insuring current activities against regulatory and business continuity risks.

As an airline owner, even if you grant that these problems are fundamentally related in the conventional causal understanding of climate change[9], namely that emissions lead to climate change, which in turn, leads to a greater incidence of extreme weather. However, it is currently impractical to act as though mitigating emissions will have any bearing on weather-related risks. The atmosphere is a commons, such that most actions taken unilaterally will only have a symbolic impact, as the power to intervene directly is distributed across many different stakeholders.

These two distributed risk issues demonstrate the challenge of managing commons resources under dispersed control. How is it that any kind of mitigation infrastructure could ever be deployed to handle these sorts of issues? One way to understand the deployment

[9] I've had the pleasure of knowing intelligent individuals who held well-informed critiques of conventional views on climate change, where they were in no position to be rewarded for their positions, and often quite the reverse. The question is not whether they are right or wrong, as the most serious claims focus around hidden common causes and are untestable except by experiments at full-scale, which are unthinkable. Instead, the right question is what exactly their claims are and how they integrate into the overall body of knowledge about the problem.

of new technological infrastructure at any scale is as a development pipeline (see Table 4). In this pipeline, publicly-funded general research is transitioned to infrastructure with privately run components, in which the risk becomes progressively more privately undertaken and normalized within an ordinary liability framework, until which time facilities need to be decommissioned, at which time earlier profits, as well as the profits from new infrastructure, can see to their maintenance. In this way, publicly-funded research and site monitoring are financed by the value generated from the productive period of similar infrastructures. The challenge of distributed risk problems is 'merely' financing a mitigation infrastructure portfolio using value captured from those activities that generate uncertain risks, and we can 'merely' sample the stakeholders of a given mitigation technology to assess its likely adoption.

Table 4: Infrastructure Development Pipeline

Early Stage Research
Objective: Must obtain grants from research sponsors and approval from peers through peer-reviewed publication
Funding: public or philanthropic (whether through a specific company or individual, industry association, or charitable organization)
Kind of Risk: public and professional (academic reputation, catastrophic experiments)
Liability: held by institution of membership, sponsoring institution, and by individual researchers

Pilot Studies
Objective: must demonstrate technical competence in an area of need
Funding: Ministry-specific public scientific funding, very early stage investment
Kind of Risk: Public (limited experimental danger and political risk)
Liability: still held by the public, likely needs approval from regional risk management officials

Incentive-driven Deployment
Objective: build the business and technical pipeline which will come to serve the needs of customers
Funding: Largely private,business development public funding
Kind of Risk: Public and private
Liability: Publicly and privately shared

Commercial Operation
Objective: Run the infrastructure as a successful business, fulfilling its role in the market
Funding: Private, should be self sustaining
Kind of Risk: Private
Liability: Privately held, and privately insurable

Maintenance and Optimization
Objective: Continue to operate effectively as a utility or public site
Funding: Public (monitoring body)
Kind of Risk: Public (failures in maintenance)
Liability: Publicly held, unless earlier negligence clearly demonstrated

Decommissioning
Objective: Transition resources to other roles
Funding: Public
Kind of Risk: Public (misappraisal of reduced risk)
Liability: Public

This pipeline suggests certain parties that will be involved no matter the underlying problem. Each stage includes primary participants, the regulators of that participation, and activists that act as meta-regulators (driving or squelching the overall process). Overall, we should expect that a new infrastructure technology will involve researchers and inventors of new infrastructure technologies; engineers and technical reviewers determining if the infrastructure is appropriate; investors and finance; clients and market makers; finance regulation; operators and facility staff; facility regulators, including environmental regulators; insurers and insurance regulators; and activists and other meta-regulators. While not definitive, this list can be checked against the methods provided later for evaluating the completeness of stakeholder coverage.

Core of the Project

This section describes a method for representing, simulating, modeling, and eliciting the stakeholder worldviews in a way which mitigates risk governance deficits and elides support-theoretic risks to political judgment. First, we will provide an overview that will explain these methods and their advantages. Next, we will describe exactly the structural and perceptual distinctions we hope to capture. Then, we will show how to capture and assemble these distinctions into risk models. Finally, we will give an elicitation procedure for producing these models.

The Complete Process: An Overview

Let us start with the big picture by considering the complete process (as shown in Figure 3) and its advantages over methods that don't include similar procedures. This process can begin at any time, and should begin as early as possible. In the beginning is a research stage, which addresses or attempts to discover any undertaking which might have a widespread physical effect, whether or not it is currently known to be risky. This document says very little about how to undertake such research processes, which may include environmental scanning, traditionally-undertaken intelligence work, or automated text analysis. This research may lead to an initial coding that establishes a baseline comparison against what is learned through interviewing. Immediately after learning a very rough constellation of stakeholders and body of general concerns, the interviewing begins. This is one point where this approach diverges from many other methods. While survey methods attempt to answer specific questions using classical confidence intervals, this method uses non-parametric Bayesian methods to peer into an open universe. Unlike open-ended interviewing, which is often constrained to the general impressions of a particular experience or service, this interview approach can discover unknown aspects of the problem domain. Further, this method then forms a more inductively powerful model than a free-form response.

After each interview, one then attempts to code the results. While the interview process is tailored to discover complete elements, it is in this stage of analysis that one finds the structure the interview has discovered. On the positive side, new patterns, overlaps, and concurrent interactions can be found. On the other hand, we will see failures to make connections, such that elements mentioned in one circumstance might never mentioned in other circumstances where they might be equally applicable. Overall, coding allows us to draw finer distinctions between the relationships between various elements than methods that do not have a formal component.

With the results of the coding, one can then undertake making inferences about the models. This inference lets us talk about how far we are in the discovery process, by talking about the rate that we discover new model elements and their features. It also lets us find differences between stakeholders, by observing differences in the models that they tend to form. These differences may highlight where strategic interaction will fail due to different underlying assumptions. Methods that don't build models of inference will need other processes to find what remains to be discovered, without which

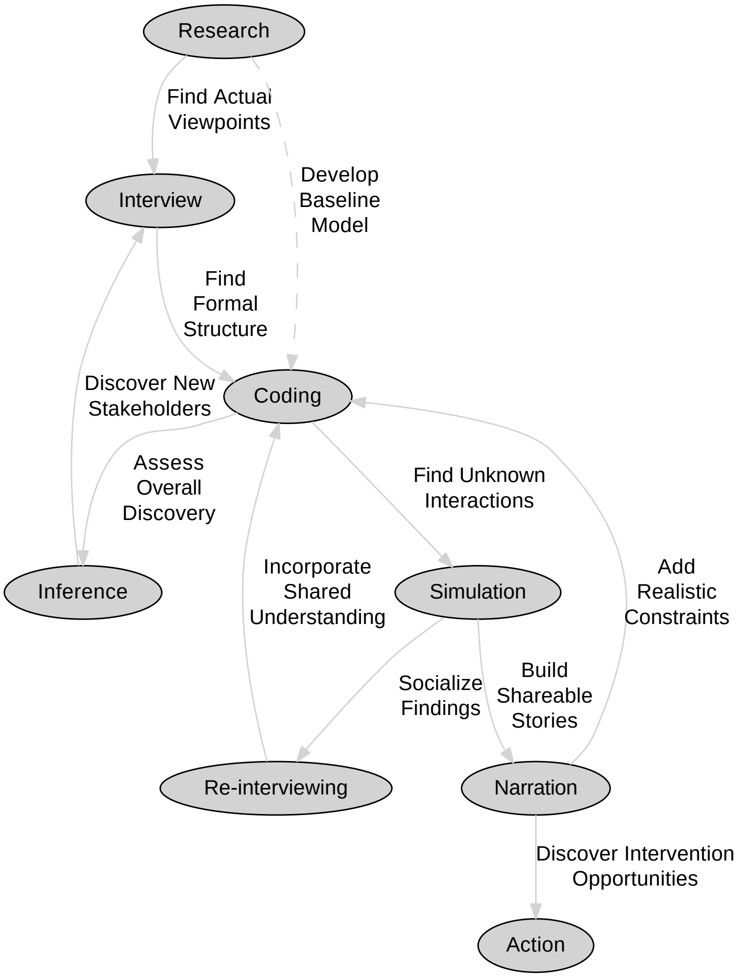


Figure 3: A Process Diagram for these Scenario Modeling Practices

they abdicate responsibility for their coverage. Models of inference also help discover variance between the facts reported about different phenomena, and variances in what different groups report about different phenomena, without which we can say very little about the overall state of knowledge.

Once a partial model is developed, one can begin to learn about potential implications through simulation. Although the participants may know about all of the phenomena and interests at play, that does not imply that they know about all of the consequences and interactions that knowledge implies. Even with an incomplete model, one can discover new paths of events. Approaches that do not undertake simulation or other methods of elaboration may not discover these interactions.

From each simulation, one should attempt to find a narrative that fits it. This narration acts as a sanity check to the modeling practices. If the process finds models that don't make any sense, this is an indication that further development of the model (or underlying formalism) is necessary[10]. Modeling processes that fail to check that their outputs are reasonable may lose their fidelity to the real world. This project is not yet prepared to offer guidance for the narration process, and leaves this kind of "sanity checking" to the practitioner. However, it is important to underline that this is a very important step, without which the process is sorely incomplete, and offering narration guidelines is a worthy undertaking for future work.

Finally, one can use these results for action. These actions include socializing the findings through a re-interviewing process or directly intervening in the processes discovered. It is useful to maintain a separation between those responsible for direct interventions and those responsible for continuing to discover new aspects to the problems, so that the findings of this processes are not distorted by an attempt to represent their stakes in a particular light.

Overall, this process may be undertaken continuously in an online fashion. In particular, we can talk about the rate at which we are discovering previously unknown impacts, and the distribution of those impacts, and therefore focus resources into discovery processes, versus other operational priorities, in a principled way. Although design is characterized as an early stage activity (which is appropriate), this process can begin at any time and should continue to be carried out as the cost of undiscovered risks is the most pressing improvement that can be made with available resources.

[10] Of course, failing to find a story from a sequence of simulated events can also be a sign that the narrator is not recognizing an unusual, but possible, scenario.

Structural and Perceptual Distinctions

We are interested in building models composed of participant-provided entities, but in order to build scenarios, we have to impose some inductive structure. The protocol is designed to elicit distinctions of:

- **the structure of the stakeholder's objective understanding**: Stakeholders will have different structures of objective knowledge. First of all, *the scope of their understanding about different processes will be different*. For example, a concrete process engineer will have a rigorous understanding of the process of making concrete, including its costs and alternatives, but may have a limited grasp on the impacts of climate change. Similarly, a climate scientist may have a solid understanding of the ecosystem impacts of climate change, but may not understand the economic challenges created by shifting between technologies with different performance characteristics. Next, *their understanding about the correspondences between events will be different*. The same observed phenomena might have different causes, indicating that different processes are at work, and that different actions are appropriate. Even if the underlying current condition is agreed upon, the consequences of further actions and events may be disputed. Also, *they may disagree in their assessment of how the nature of these correspondences could be altered by the dynamics at which phenomena interact*. They may believe that particular phenomena will unfold with different rates or intensities of change. Even if stakeholders agree about the potential correspondences between phenomena, *their assessment of the likelihoods of these correspondences holding may be different*. A stakeholder may think one consequence is much more likely than another, or that some occurrence will indicate one cause much more likely instead of another. Finally, although stakeholders may agree on underlying processes in every respect, they may *disagree on the salience of observable factors* in determining the condition of the underlying process. In summary, *the mental models of stakeholders* *might disagree in objective content, in causal connectivity, in dynamic interaction, in magnitude of correlation, and in the degree of observability*.

- **the structure of the stakeholder's subjective perceptions**: Stakeholders will also have different subjective perceptions. First of all, each stakeholder will *identify different kinds of losses they may endure*, including loss of life, loved ones, livelihood, property, resources, comfort, opportunity, and knowledge in the security of natural world and future generations. Next, the different stakeholders will *experience different magnitude of these losses* in different situations. Also, when faced with many different kinds and magnitudes of losses, stakeholders may choose different trade-offs between mixtures of losses. Finally, each stakeholder will *assess the different kinds, magnitudes, and trade-offs of other stakeholders differently*. What one stakeholder will value, another may not. However, at the same time, there is a difference between what one imagines the impact of a loss will be like, so the assessment of other stakeholders may not be a misassessment. Altogether, *the assessment of loss of stakeholders may be different by kind, magnitude, trade, and observer*.

- **the stakeholder's understanding of objective orientation**: Given the stakeholder's understanding of the objective structures and subjective concerns for the matter at hand, there is still their perception of what the current situation is. This includes both the current condition, but also any ongoing trends, processes, or activities. This may also imply a perception of benefits or losses being conferred upon stakeholders currently. The understanding of the current situation includes what is believed about the past.

- **subjective perception of objective knowledge**: Stakeholders may also have different appraisal of the objective understanding of other stakeholders. This works both positively and negatively, as they might believe that other stakeholders have better or worse understandings of particular knowledge areas, grasps on the factors determining causal outcomes, and expectations of likeliness. The awareness that others know more may occasionally serve as an indication of trust, and often an expectation of responsibility for management or regulation. Similarly, the suspicion that another stakeholder does not understand an area with an impact upon other stakeholders may be taken as an impression of negligence.

Assembling Structure into Cohesive Risk Models

Given our commitment to model the distinctions above, we would like to model input from the stakeholders such that we capture that input with fidelity, and yet do not introduce any background information beyond those distinctions. At the same time, we wish to not irritate our stakeholders by asking about common knowledge, nor to produce models that are incomprehensibly large, but largely composed of trivialities. This produces a very challenging representation problem, in that we want to capture the right features for risk models while maintaining usability. This section first creates modeling primitives to meet this objective. Then, we demonstrate how these model components come together in order to simulate potential outcomes. Finally, we show how this model comes to be built in elicitation procedures by giving a statistical characterization of its discovery process.

Structural Elements

In this section, we introduce the structures of our scenario model. Here, we are aiming to create a scheme for representation that can be suitably complete and precise, yet realistic and flexible. In order to explain this model specifically, the following employs notations from mathematics (see Appendix D for a guide to the mathematical notation used within this paper), although hopefully the explanations will suffice. Along with the mathematics, we also present a programming language for representing this model that demonstrates how to put these structures to work (a complete grammar of this programming language can be found in Appendix A). Given this introduction, let us begin.

The first thing that needs to be established is the objective content of the observer: what is it they are talking about? A structure is a description of physical facts and the relationships between them. Let us call a particular example of such a description s, and that there is a set of all possible structures, S, such that $s \in S$. Without loss of generality, we can confine the description of a structure to a given point or interval of time[11], t, if we can say for every component and relationship was present (or absent) throughout[12] this time. Let us denote this confinement as $s[t]$. However, what we are trying to capture is components or behavior that are static with respect to the phenomenon under analysis, so structures should not generally include changes of conditions. However, structures can represent unchanging rates or norms, even though those imply activity; for example, a steady flow or an average day of sales are perfectly fine to include in a structure. In general, it may be very difficult to tell if the descriptions of structures are contradictory or not. For example, if a house is white, but somebody else says it is brown, it may be a brown and white house, or we could have conflicting accounts.

There are many ways these structures can be modeled. The absolute minimum we could do would be to list out all the participant said when elicited about the situation. Let us say first that a structure could be a bag of descriptions. We have some number of short textual labels, or tags, that describe a particular condition, and that each of these tags is one of the overall set of tags used in the model ($tags \in Tags$). It could be that these tags designate quantities, percentages, or are only true in some fuzzy sense, so it is useful to associate each of these tags with a real-valued weight ($w \in \Re$), such that each additional point of data is taken to be an ordered pair $d = (tag, w) \in (Tags, \Re) = D$. By default, the tag value is 1, which we interpret to mean that the proposition suggested

[11] As a rough approximation, interval temporal logic [Allen and Ferguson, 1994] will serve as a guide.

[12] Optionally, during semantics can also be specified, but they are not the default.

by the tag is true. Therefore, a structure contains weighted tags.

This is fine, as far as it goes, and it can go far, but there will often be more structure to the structure, so to speak, than a flat list. Therefore, it makes sense for structures to contain other structures, and so we can arbitrarily nest structures, and define structures recursively as $s \in S = D \bigcup \mathcal{P}(S)$. In summary, structures are nested, weighted tag clouds. Although it will not be shown here, this is sufficient to represent an arbitrary set of relations between boolean and real-valued quantities.

Now that we have introduced structures, let us show how to represent them. Here's an example, where we are saying that the United States's Great Plains region is at the risk of transitioning to a desert ecosystem, and that this is currently the case according to a Canadian carbon air capture technology expert:

```
structure usGreatPlainsAtDesertRisk
  (usGreatPlains atDesertRisk)
  current according to CanadianCACTechExpert.
```

We see that first we declare a structure with the name 'usGreatPlainsAtDesertRisk' and that this structure consists of a cloud with two tags (with default weight one). Giving the cloud is optional, in which case the structure is taken to be the cloud with one tag, the name. Anytime we declare that a particular condition is, was, or will be the case, it becomes necessary to identify which stakeholder asserted that it was so.

It is useful to talk about structural expressions, $se \in Se$, which are equivalence classes defined by logical expressions over structures. For example, we might be interested in all structures with the tag 'atDesertRisk'. We could say that as `contains($region, atDesertRisk)`, such that `CanadianCACTechExpert` would say that $\$region$ could currently be satisfied by `(usGreatPlains atDesertRisk)`.

Structures are always partial descriptions, so multiple structures are used to describe the overall condition at the same time. The complete set of structures at any given time describes the state-of-affairs, or state for short, which itself is a member in all possible states ($x \in X$). Every state is merely the structures it is composed of ($x \subset S$), and states admit temporal confinement in the same way as structures, such that we can talk about $x[t]$ in the same way as $s[t]$.

It's very useful to distinguish some structural elements from others. Prime among these are stakeholders, $sk \in Sk \subset S$. Stakeholders indicate what they perceive to be the case, but they also can participate in those perceptions. Stakeholders who are providing the information are called the 'observing stakeholders', $ob \in OB \subset Sk$. Here are two stakeholders that we have seen before:

```
stakeholder UnitedStates.
stakeholder CanadianCACTechExpert.
```

Stakeholders are structures for a very important reason: stakeholders are not simply labels, but instead can represent composites of different interests based upon varying conditions. A farmer who loses their farm to a flood likely has a different set of concerns about agricultural policies afterwords.

Stakeholders are informed of states-of-affairs through observations, or rather observable factors, which are structures composing the state-of-affairs in their own right, such that $o \in O \subset S$. Observations can be declared in a straightforward way. Each stakeholder will report some subset of the observations as being the case at some particular time, $O_{ob}[t] \subset O$. Here is an example of an observation:

```
observation drought.
```

Every underlying set of states can potentially generate some set of observable factors, according to some distribution $\theta \in \Theta : \mathcal{P}(S) \rightarrow P(\mathcal{P}(O))$. These distributions can be subject to various dependency structures, but can also be expressed in an uncomplicated way when a stakeholder expresses them as such. As different observation functions, or sensings, are provided by different observers, we designate the observation functions elicited from a particular observer as $\Theta_{ob} \subset \Theta$. Here is an example of an observation function:

```
observe (drought) when (climateChangeCurrent)
        according to CanadianCACTechExpert.
```

As we have direct access to observables, but not the underlying conditions that caused them, it is often appropriate to talk in terms of the inverse of θ, or $\theta^{-1} \in \Theta^{-1} : \mathcal{P}(O) \rightarrow P(\mathcal{P}(S))$.

It may be important to consider whether stakeholders will actually observe these factors, but to 'notice' is just one of many actions a stakeholder may undertake. As such, we generally designate actions as another kind of structure $a \in A \subset S$.

```
action doNothing.
```

Given that we can represent actions, structures, and states, we are now prepared to understand how actions initiate change in the state through events that affect structures. Events change one structure into another over a period of time according to some probability distribution, such that $E : X \rightarrow P(X \times T)$. This is accomplished in terms of its structures, such that each event only changes some portions of some of the structures, leaving the rest unchanged. So, we can say that each event will change structures in the state that match some structure expressions into structures that match other structural expressions, or $E : \mathcal{P}(Se) \rightarrow P(\mathcal{P}(Se) \times T)$. This implies that, for all the states-of-affairs that might be affected by the event in the same way, and all of the substructures which satisfy the expressions in the same way, we only need to specify a given event once. A given observer will report some subset of events, which we designate as $E_{ob} \subset E$. For a given event ($e \in E$), let us describe the set of structural expressions forming the precondition as $Pre(e) \subset \mathcal{P}(Se)$, and postconditions as $Post(e) \subset \mathcal{P}(Se)$. Symmetrically, we can talk about the subsets of events that have a given expression ($se \in Se$) as a precondition $PreSet(se) \in E$ or postcondition $PostSet(se) \in E$. Let

$p_{e,t}$ be the distribution of the duration of event e, given that it does occur. Here is an example of an event that leaves both probability and timescale unspecified:

```
event climateChangeRegionalImpacts
      if $climate contains [climateChange, unmitigated]
         and $region contains [atDesertRisk]
        and $region does [doNothing]
      then $region becomes (Desertification) from (atDesertRisk)
      according to CanadianCACTechExpert.
```

This example states "if climate change continues unmitigated, should those regions at desert risk take no measures otherwise, then the region will experience desertification." Notice the use of 'does', which is a special form of structural expression for handling the actions of stakeholders.

As a technical concern, it can be useful to divide an event into two events, from the initial structure to an intermediate structure, and from that intermediate structure to a final structure. This prevents the structure from triggering other events inappropriately while in progress. It should also be noted that while processes are often cyclic and concurrent, that does not preclude us from representing them through point-wise events, such that the overall result of a process on the structure in a given time period is merely the composition of events between those periods.

However, there is one more technical concern that does need to be handled in order for events to make sense. Events are often dependent upon each other, and sometimes completely so. Suppose a student takes an exam. This may lead to the student passing the exam, or failing the exam, but not both. These two events are mutually exclusive, as are `carbonTechTestFail` and `carbonTechTestSucceed`.

```
depending on carbonTechTestFail
          mutually exclusive carbonTechTestsSucceed
          according to CanadianCACTechExpert.
```

In general, dependences mean that if some condition occurs, then the likelihood of another condition is then modified. We can say that these dependencies are $d \in D, D : \mathcal{P}(Dt) \rightarrow P(\mathcal{P}(Dt))$, where $dt \in Dt$ is the set of dependency terms, or expressions that can be used in dependences. Events are dependency terms ($E \subset Dt$), as are sensings ($\Theta \subset Dt$) and the yet-to-be-introduced anticipations ($An \subset Dt$). We can also designate subsets of dependences that have terms of these various types ($D_E \subset D, D_\Theta \subset D, D_{An} \subset D$, respectively). In general, if a distribution is associated with an element, it is also appropriate to be able to specify conditional distributions for that element. Dependences are generally unidirectional unless designated to be mutual or as independent (which is necessarily mutual). Similar to how events have preconditions and postconditions, we can refer to the antecedent ($dt \in Ante(d) \subset Dt$) and consequents ($dt \in Cons(d) \subset Dt$) of a dependency. Symmetrically, we can talk about the set of dependencies that are dependent upon a given antecedent, $UponSet(dt) \subset D$, or those that have a given term as a dependent consequent, $DepSet(dt) \subset D$.

Now that we can represent the structure in the world, and have the dynamics to put it into motion, let us now look into the pragmatics of those structures, or why stakeholders are motivated to act in the ways that they do.

Each stakeholder has some set of criteria $c \in C$ to which they respond, and which they are motivated by. These criteria are incommensurate, and as such different stakeholders will trade-off between these criteria differently at different times, if they find themselves able to make a trade-off at all, to the extent that any of these criteria, improved at the cost of another, leads to regret. With each criteria, one also specifies the way it should be assessed, whether to maximize, minimize, or to pursue a certain value. For every criteria there is also a margin of indifference, ϵ_c, for which every stakeholder is indifferent to changes between, and a discount parameter α_c, which will be explained later.

```
criteria agricultureProduction maximize.
```

Agricultural production is not a good criteria, because it does not indicate a direct harm felt by any particular stakeholder. We can understand it to be proxy for farmers having to leave their farms, for individuals going hungry for being unable to afford food, and so on. On the other hand, it might mean that individuals are becoming employed in other jobs, and more food is being imported through trade. On the other hand, the stakeholder from which this information is being elicited may very well think of this as a good in itself, having never contemplated that changes in agriculture production is a proxy for a whole host of goods and harms, and may not be prepared to think of them in this way.

Each stakeholder is impacted by a different set of goods and harms in different ways. We represent the total impact of any given structure, to any particular stakeholder, for any particular criteria, as a reward function yielding a real number $R : S \times Sk \times C \rightarrow \Re$. Impacts can also be specified relative to the scale of the weights in the structure, which implies that impacts can also be supplied qualitatively. A given observer will report some subset of the overall impacts: $R_{ob} \subset R$. It is important to note that each observer does not only report their own overall impact, nor is each observer's reported impact taken as the final word. Additionally, let us say that R_c is the set of rewards that yields a particular criteria, and that $value(r)$ is the reward or loss assessed ($\Re$).

```
impact cropLoss
       high declining agricultureProduction
       to agricultureRegion if (drought)
       according to CanadianCACTechExpert.
```

At this point, it is worthwhile calling attention to some of the features of this model. First of all, it can represent incommensurate forms of evaluation. The same state can cause a stakeholder both benefits and harms, by different criteria, such that both acting to cause the state and acting to avoid it cause regret. More novel, but of equal importance, is the fact that the stakeholder is constituted of structures, which means that the same stakeholder can experience different rewards and harms based upon their current condition. Although the ability of actions to change desires or preferences is not typically explicitly represented in computational planning approaches, to do so is the basic operation of analytical sociology [Hedstrom, 2005], and is essential to these models being sociologically plausible. There are ways the stake of a stakeholder can change. They may be persuaded, or the environment of the stakeholder can change, in response to which the stakeholder adapts their preferences, in a kind of cognitive dissonance. The preferences of a stakeholder may also change due to changes in their material stakes; for example, a farmer who has had their farm repossessed by a bank would certainly have less reason to care about agricultural policies.

Stakeholders will often anticipate that they or others will respond particular ways in a given situation. We model these anticipations as saying that if the current conditions match some structural expressions, then particular stakeholders will undertake actions with some distribution, sometimes with respect to particular structures identified by a structural expression, or more formally $an \in An, An : X \rightarrow P(Sk, A, Se)$. Here is an example of such an anticipation:

```
anticipate usDoesNothing stakeholder UnitedStates will doNothing
           according to CanadianCACTechExpert.
```

Finally, let us say that all of the elements that we have described so far, are model elements, ME, portions of an overall model. For any of these, a stakeholder may choose to defer to the knowledge of a stakeholder. Similarly, a stakeholder may indicate the opposite: that when it comes to this kind of matter, the other stakeholder has no idea what they are talking about. $df \in Df, Df : ME \times Sk \times Sk \rightarrow [0, 1]$, where $[0, 1]$ is set of real numbers from zero to one, inclusive. This number is the degree of deference that the first stakeholder will grant the second stakeholder. $Df_{ob} \subset Df$ is the subset of deferences granted or reserved by a particular stakeholder.

Together, structures, states-of-affairs, stakeholders, rewards, criteria, events, dependences, actions, observations, sensings, anticipations, and deferences are the structural elements we need to represent an object-level causal model of multiple impacts, M (such that $M = (S, X, Sk, R, C, E, D, A, O, \Theta, An, Df)$). We can see that M is a causal model, with S playing the role of variables, and E, An, Θ, and Df playing the role of functions.

The Challenge of Loss Over Time

Using the tools of the previous section, we can now describe any number of situations, but we have not described how these elements interact with each other and lead to situations unfolding into each other. This section hopes to rectify this deficiency by describing how the elements in the last section come together in a simulable way. This section focuses on some of the challenges in appropriately framing how to evaluate concerns over (potentially intergenerational) periods of time, while Appendix E does the technical work of assembling these elements into time-series and giving a more mathematical account of the simulation process.

How do we think about the rewards and losses to stakeholders over time? One way is to consider all of the losses over time, either in sum or averaged across time. This formulation has a quirky behavior: if we sum them, then recurring rewards and losses go to infinity, while if we average them, one-time gains and losses disappear entirely.

The sum of rewards and losses is fine for an individual stakeholder, who is objectively finite, but if we want to think about ongoing populations of similar stakeholders, then these mathematical complications render assessment difficult to interpret. Despite those complications, the average quantity has some desirable properties. For example, suppose that a renewable resource is destroyed. In this case, all future generations may be deprived from its use, and an infinite amount of potential reward was squandered. It may be that this average method of evaluating costs has virtues in certain resource management problems.

Unfortunately, the average cost formulation has severe problems in the context of our formalism. First of all, if the individual stakeholders have limited lifetimes, then no particular stakeholder interest is represented by this formulation. Even worse, the average stakeholder interest is arbitrarily far away in time from that of any current stakeholder, so we would expect any change would become compounded, leaving this estimate useless for many. It also shows a fair amount of arrogance to presume that the objective state-of-affairs is going to be similar enough, and predictable enough, to be evaluated. Although we would like to represent the cross-time interests of potential stakeholders, these complications suggest it is better to assume that we can only anticipate our conditions and their consequences in a more limited way, that decreases the further we move from our current time. Let us say that discounting is the process for decreasing the evaluated rewards and losses in proportion to their distance in time, and that the rate of this decrease is done according to some discount parameter (represented here as α).

This discount parameter has a number of pitfalls that one should watch out for. One common mistake is to forget that the discount rate should be different depending upon the kind of harm it is arbitrating [Sunstein, 2007]. It is always true that we would prefer that harms occur further into the future as opposed to the present time. Therefore, it is only sensible that we would be willing to pay slightly more to avoid a harm that will manifest itself soon as opposed to one that will occur at some time away. However, this does not imply that the difference will correspond to existing interest rates. Therefore, when doing a financial analysis comparing the cost-benefit analysis of various programs, it is inappropriate to depreciate other risks at the same rate as financial assets. Therefore, discount rates are criteria specific (α_c).

Discount rates can also be problematic for approaches that mitigate long-lasting phenomena. If we were to undertake an economic analysis of carbon sequestration, then one would think that it is appropriate to account for the expected period of containment for the method under study [Herzog et al., 2003]. However, a shorter discount rate, including the expected lifespan of everyone currently living, may also be deemed appropriate, pushing the assessment from an economic issue to one of intergenerational fairness.

Given these caveats, we can now describe a simulation process that evaluates how individuals are impacted by given conditions, how they act given those conditions, and the conditions that result from the events caused by their actions.

Algorithm 1 Scenario Model Simulation Run

```
Sample initial conditions
while Possible absolute discounted risk is greater than some small quantity of indifference for any criteria do
    Evaluate the rewards and losses for each stakeholder
    Sample the actions that stakeholders will take
    Sample the events that result from the current state and those actions
    Based on those events, determine the resulting state-of-affairs
end while
return Everything that happened in the simulation
```

When we sample, we are evaluating probability distributions while respecting dependencies. Dependencies are handled by first resolving all terms that have no dependencies, and then resolving their dependents. When all terms are resolved except those that are mutually dependent, the cycle of mutual dependence is broken by resolving terms at random until at least one term has its ancestors resolved.

This simulation procedure is likely the simplest simulation procedure that will work for this model, but for now we leave writing a simulator that supports different event durations and similar improvements for future work.

Assembling Pragmatic Causal Categories

Although we have gone quite far, we have not yet covered an important aspect of this work, namely how we can tell where we stand in the discovery process. It turns out that there are some fundamental questions that have to be answered to address this question.

For example, what does it mean for the interests of two stakeholders to be the same? Consider two neighboring farmers who have been friends for a long time. If one of their farms experiences a catastrophe, say a long period of flooding that renders their land underwater, and thus unfarmable, then both will experience negative consequences. The farmer with the flooded land will experience financial ruin and the associated challenges with losing part of one's home, while the farmer who has been spared will experience the challenges associated with a friend making a difficult transition. These farmers have identical concerns, even though it makes a great deal of difference to them who's farm is underwater, in that discovering the concerns of one is enough to posit the concerns of the other. We can also say that these concerns sufficiently similar even if there are some salient differences between them, for example if one of the farmers has a bit more savings, then the consequences of a flood might not be as immediately dire, but their concerns would be still be similar. However, if one of the farmers was independently wealthy or actually primarily employed otherwise, then these farmers may actually have substantially different concerns.

We can say that a given stakeholder (sk_1) shares a situational preference similar to another stakeholder (sk_2), for a given criteria c, if sk_2 prefers some structure s_1 to s_2 by that criteria, then sk_1 either has the same preference, or at the very least dislikes it by some margin of indifference (ϵ_c).

$$\begin{aligned} sharePref(sk_1, sk_2, s_1, s_2, c) = \quad & R(s_1, sk_1, c) - R(s_2, sk_2, c) > 0 \\ \implies & R(s_1, sk_2, c) - R(s_2, sk_1, c) + \epsilon_c > 0 \end{aligned}$$

This is trivially true if sk_2 has no such preference, so we say that in order for two stakeholders to have a similar preference, if either of them has a preference, then the other shares it or is at least marginally indifferent.

$$\begin{aligned} simPref(sk_1, sk_2, s_1, s_2, c) &= sharePref(sk_1, sk_2, s_1, s_2, c) \\ &\wedge sharePref(sk_2, sk_1, s_1, s_2, c) \end{aligned}$$

We can say that two stakeholders have the similar concerns if and only if, for all the concerns under consideration, and comparing all situations, they have similar preferences.

$sk_1 \sim_C sk_2 \iff \forall c \in C, \forall s_1 \in S, \forall s_2 \in S, simPref(sk_1, sk_2, s_1, s_2, c)$

This is a useful intuition to have for understanding how to assess if we have discovered the stakeholders, but there are three problems worth understanding.

For one thing, although these farmers have similar concerns, they may not be similar stakeholders, for we are also concerned with differences in their capabilities to mitigate risks. For example, if one farmer had, due to differences in the physical geography of the farms, the capability to install an effective drainage system, while the other farmer did not, then these farmers would still be substantially different.

Secondly, from a risk governance perspective, this definition may be overdoing it. When doing the analysis of a given risk situation, we are often not as interested in risks to specific individuals per se, as compared to individuals in particular roles. For example, if a given company is designing a coal-fired power plant, then although the designers of the plant may very much like Bruce, who will be employed as an operator mechanic in fuel prep, they are merely analyzing the risk to him due to his role as an operator mechanic, and not to the risks undertaken in other aspects of his life, which we presume are either undertaken solely at his discretion or under the regulatory responsibility of others. The methods described so far are, for this reason, already cut across categories. However, what they do not do is to reassemble the full risks or capabilities of any particular class of actor, but only the apparently relevant situational risks.

Finally, and most importantly, it is simply unrealistic that any kind of elicitation will pin down all of the concerns of a stakeholder, instead of the ones that are merely currently salient to the situations that happen to come to mind. The definition is obviously an idealization of what we could learn.

What it is practical to learn are approximate categories and rough attribute sets corresponding to the elements that we are eliciting. We know a stakeholder is similar enough if, in the same kinds of situations, they experience the same kinds of impacts. We can now put the non-parametric statistics we have developed earlier to work. The mathematical version of this model can be found in Appendix F.

First of all, we would hope to discover approximate categories of stakeholders, the kinds of general situational structures they will encounter, the different reward and loss criterion that they have, and the sorts of actions they might take. There could be an infinite number of each of these, but will be found in different percentages in the population with diminishing returns, so let us say they are distributed according to the CRP. The degree to which we expect to discover new categories for each of these is interesting, as they determine the overall level of exploration, and we will revisit them. For now, we just say that stakeholder categories, structure categories, criteria categories, and action categories are modeled with a CRP distribution.

Although each structure category may vary in terms of the tags that describe it[13], we expect that some will be used more frequently than others, again with diminishing returns, and thus we can say that the tag distribution with respect to a given structure category has an IBP distribution with a small exploration parameter, reflecting the fact that we expect the most salient details to be applied readily and more subtle insights to be rare.

Does a stakeholder experience an impact relative to a concern in a particular stage? We can say whether or not a stakeholder category tends to have a particular criteria for a

[13] We know that structures contain more, well, structure, than is captured by a tag set. We will save a model that probabilistically builds predicates for future work.

particular structural category, which we would expect to have a beta distribution with a small gamma (say 0.1), indicating that it should be very likely or very unlikely that a stakeholder category should have a particular stake given particular conditions, that being the definition of a stakeholder category.

For simplicity, we can treat the magnitude of concern[14] as a multinomial over the discrete categories "strong reward", "weak reward", "no reward, no loss", "weak loss", and "strong loss". We will use a uniform prior.

We can say that a state-of-affairs consists of all of the structures that are currently the case within it. We expect to receive different lists of structures each time we elicit, but again that the variation of structures we find has diminishing returns. Therefore, states-of-affairs could reasonably be represented as a IBP mixture model over structures, generated from an underlying CRP mixture model.

How should the expected state-of-affairs discovery rate and the structural-diversity of states-of-affairs be chosen? If set smaller, it assumes that structures are more likely to be concurrently the case, while if larger, it presumes that different structures are more likely to correspond to different states of affairs. As it stands, we expect both that the states-of-affairs are highly-overlapped, but also that when the stakeholder is describing different outcomes, these descriptions are largely similar except for key factors[15] Overall, it isn't clear at this time what this implies, and we recommend observing the diversity of elicited structures to find appropriate parameterizations.

The multiplicity of structures in states-of-affairs is not the only time we will want to talk about combinations of elements occurring together. For example, we might want to say some set of events is dependent upon another set of events. For this reason, let us also refer to categories within states-of-affairs as categories of structural combinations.

[14] You may wonder why we would represent "no reward, no loss" as a magnitude of concern, and the answer is that it is a matter of relative comparison to the conditions connected by events; consider the phrase "stop hitting me" for example.

[15] Indeed, this similarity between possible worlds is a popular philosophical framework for counterfactuals [Lewis, 1973], although it is prior to the intervention-based conception used here.

Let us say that sets of events categories occur together in such a way that they can be sampled as an IBP mixture model with a CRP prior.

Given that we can talk about states-of-affairs, we would like to be able to talk about which actions are possible and relevant for stakeholders in those states-of-affairs, no matter how likely they are. We can say that such an action is specified as being salient, event if deemed impossibly unlikely, if the relationship is indicated. The distribution over possible anticipations is beta.

If a given action is possible, how likely is to to be undertaken? We would say that it is likely that if the observer knows how to specify it as a possibility, they also likely have a suspicion of whether or not it will be undertaken. For that reason, we can say that the distribution is almost uniform but slightly favoring the extremes, as in a beta distribution with parameters slightly less than one.

Of course, as established before, actions do not stand alone, but they are anticipated to have their compliments and substitutes, so there might be a dependence between sets of anticipations, where were need to specify both a likelihood of the dependence (or the indicator) and the likelihood given the dependence. These are specified as beta and uniform, respectively.

Given a categorical state-of-affairs and some combination of actions, we can ask if another state-of-affairs is the result of that event, and if so, how likely that is. These in turn are specified as beta and uniform, respectively.

Event categories may also have dependencies, parameterized by both presence of effect and their likelihood given the dependence. These are also beta and uniform, respectively. The distribution over sensings given observations and underlying structures is similar to anticipations and events is also taken to be beta and uniform, as are the dependencies between sets of sense categories.

Finally, a category of deferences means that over any set of model elements a deference may be given to a particular category of stakeholders. This deference also has a likelihood, where negative deference is to cast suspicion on the knowledge of a particular stakeholder. These too is distributed beta and uniform for occurrence and strength, respectively.

What are we to make of the expected rates for discovering new stakeholders, structures, criteria, and actions? Is it not the case that, if we take unknown harms to unknown stakeholders, in unknown conditions, that we must, as a matter of precaution, assume that these parameters are very large and that, as we appear to hit diminishing returns we are merely unlucky and need to persevere? Or is it instead the case that, in representing the interests of any significant group of the general public, we merely have to capture the most generally held values and stakes, as well as the conditions and actions that could possibly disrupt them, and that we must let marginal concerns be marginal? Whether one pursues diversity or representativeness is a topic of debate in the framing of deliberation and its purposes [Renn, 2008], so the question is very likely insoluble. However, framing this choice as picking between discovery rates allows for some trans-ideological guidelines:

1. At all times, once a design activity has been undertaken, we can assess the rate at which the number of these factors has been growing. Although it may be important to assume a rate to gage the number of initial participants, there is no reason not to reassess the most likely rate of discovery, given some examples.

2. Some design activities are much less expensive than others. For this reason, generative activities such as brainstorming should initially aim for quantity, while more costly elicitation activities may expect this to be smaller. In all cases, it is possible to overlap the results of activities to see if field results support or undermine generated results.

Finally, let us take a step back, and think about the overall model presented here. As we ask, we discover dependencies between different clusters, and this model allows us to look at changes in the structure of what we know as we ask. However, in order for the relationships in this section to hold, the processes that discover them must obey a particular constraint: the elicitation processes must be exchangeable. In other words, the individual samples underlying inference this this model must be undertaken in any order. In practical terms, this means that we should not tell a participant what another participant has said until we have first listened to what they have to say.

An Interview-based Elicitation Process

In this section we turn from structure to process. We have a domain-general model for discovering causal information, but how do discovery processes actually construct it? The strategy that we use here replicates a forward simulation of potential outcomes, asked in a depth-first fashion.

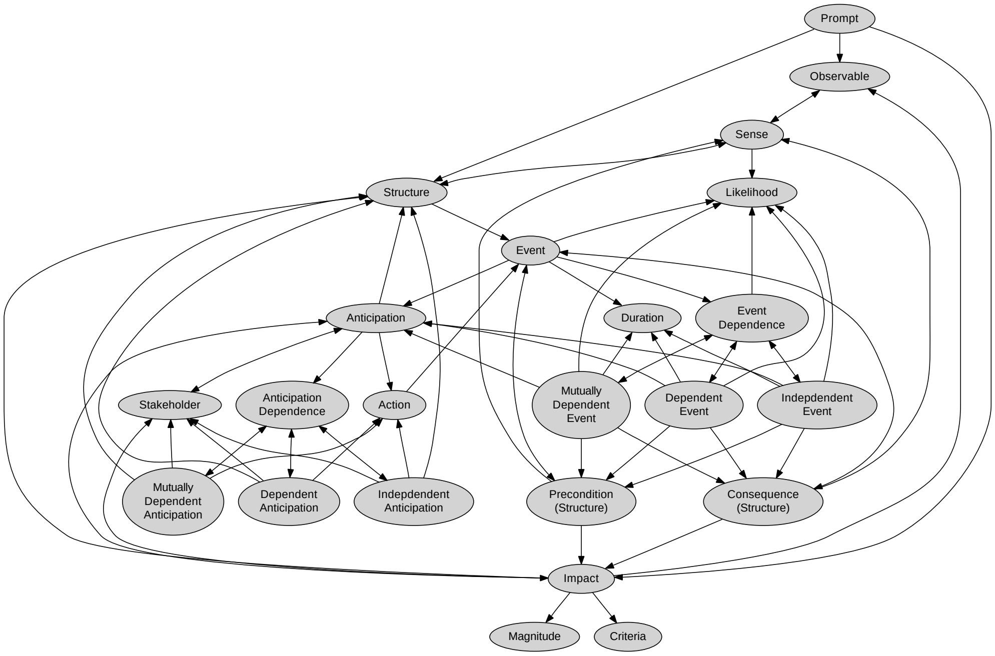
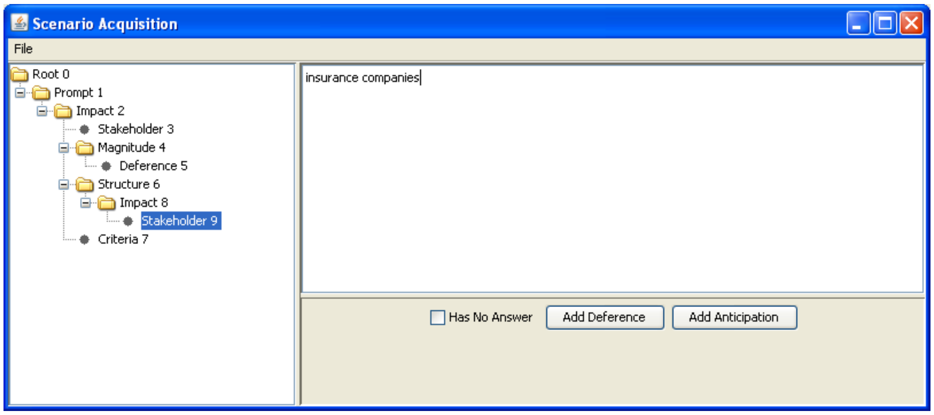


Figure 4: The Potential Paths for Questions

An interview using this approach are undertaken in segments. These segments may either be exhaustive (where every possible question in the search tree is asked) or bounded to a preagreed time. In each of these segments, first a domain-specific, but still generic, prompt question is asked. For distributed risk problems, choose separate prompt questions at the problem impact and mitigation adoption scales, to check to see where these two aspects join together for the stakeholder, if at all. The segment then proceeds by asking general questions that, in effect, simulate what what they suspect will happen, how that will affect the stakeholders involved, and how they will behave as a result. The number of paths that potential questions could be taken grows quickly and is very complex (see Figure 4)[16]. As a result, two measures are undertaken to assist in keeping track of the interview. Ideally, the interviewer can use a scenario acquisition tool that keeps track of what questions yet need to be asked (see Figure 5). This tool allows for different threads of potential events to be pursued while automatically returning to previous questions as lines of inquiry are exhausted. Although it is far easier to use such a tool, it can be undertaken by hand using an arrow-based notation, examples of which are given in Appendix B.



insurance companies

Figure 5: A Scenario Acquisition Tool

As you can see in the tool screenshot, each question is numbered. Provide participants with numbered form with empty boxes to write notes for their references. When we

[16] Note that the expectations of what others know, which I call deferences, are asked about in all stages and thus are not shown on the graph for simplicity

want to ask questions about answers earlier in the interview, the participant will have a context to answer. A complete example of a telephone script is given in Appendix C. After the end of questioning, participants should be given an opportunity to make additional comments clarifying their earlier responses.

Undertaking the path of questions in a depth-first ordering is an assumption-free traversal of the question tree that promises completeness. In an interview without time-bounds, it is guaranteed that the interviewee will be asked every line of inquiry. Furthermore, of the paths that are complete, this method promises the minimum number of digressions: every question will correspond to the most recent possible previous question. The downside to the depth-first method is that it can be a long time before alternatives that occur early in the process. As an alternative, the interview can be directed by rules that balance questions back to more recent events while preserving eventual completeness. For example, we could allow only a maximum disparity in the tree depth between topics with unanswered follow-up questions. However, the advantages of balancing methods is made at the cost of making unnecessary topic switches, which you really would not want to do if the interviewee was right in the middle of answering questions well.

Let us now look through examples of each step along the potential paths of non-prompt questions. All of the other statements necessary to undertake a prototypical interview with two prompt questions is given in Appendix C. Here, as elsewhere, we use *italics* to indicate when the participant is quoted.

Prompt → Impact

- Is anyone impacted by *participant response to prompt question*, either positively or negatively?
- Is anyone else impacted by *participant response to prompt question?*

Prompt → Structure

- What is happening right now in regard to *participant response to prompt question*?
- Is anything happening now besides *participant response to prompt question*?

Prompt → Observable

- What can we observe as a result of that happening?
- Regarding that, what could the average person observe that would indicate to them that this is the case?
- What else might we observe that indicates these conditions?

Structure → Sense

- Are there things we could observe to tell us if this is the case?
- When this happens, what will we observe that lets us know what has happened?

Event (all kinds) → Consequence

- What happens as a result of *described event*?
- Does *described event* have any other consequences?

Event (all kinds) → Precondition

- What could cause *described event*?
- Is anything else needed to cause *described event*?
- Are there any other causes for *described event*?

Event (all kinds) → Duration

- And how long will that take to occur?
- After that starts to happen, how long will that take to really get going?
- How long will it take before we see the affects of that?

Event (all kinds) → Anticipation

- And if that happens, will anybody take actions as a result?
- Might any other actions result from *described event* happening?

Impact → Stakeholder

- Who will be affected by this?

- Who will be affected this way?
- Will anyone else be affected in this way?

Impact → Criteria

- How are they harmed (for example, physically, financially, reputationally)?
- How exactly do they benefit?

Impact → Magnitude

- How many *direct quote of stakeholders impacted* are affected in this way?
- How much does this harm *direct quote of stakeholders impacted*

Impact → Anticipation

- As a result of *described impact*, how could *described stakeholder* respond?
- Are there any other ways *described stakeholder* could respond to *described impact*?

Impact → Observable

- When *described impacted stakeholder is described impact*, what might we observe?
- What are some signs that we will be able to tell that *described impacted stakeholder is experiencing described impact*?

Impact → Structure

- Under what conditions might *described stakeholder* be *described impact*?
- Are there any other conditions under which *described stakeholder* might be impacted in this way?

Structure (all kinds) → Impact

- As a result of being in that condition, would any of the stakeholders experience gains or losses?
- Are there any potential harms to being in this condition, or any rewards for that matter?

Structure (all kinds) → Sense

- Is there any signs that this has become the current condition?
- What would be some indicators that *described structure* has come to pass?

Structure (all kinds) → Event

- Could this state-of-affairs cause any other events?
- What might happen as a result of *described structure*?

Event (all varieties) → Dependence between events

- Is this more or less likely if something else happens?
- Are there any other events that could cause or prevent this from happening

Event (all varieties) → Likelihood

- How likely is that to occur?
- When you say *described event* will happen, how likely is that?
- What do you think the odds of that happening are?

Observable → Sense

- If one observes *described observable*, does that tell us anything else about the current conditions?
- Is there anything else that might be true about the underlying conditions if we observe *described observable*?

Sense → Observable

- What signs might we able to observe indicating that?
- What are some of those indicators?
- Is there anything else we might be able to observe that would allow us to infer that might be going on?

Sense → Structure

- When *described observation* supports some underlying facts about the current conditions, what are some of the conditions it indicates?
- When we see *described observation*, what to we know to be going on?

- What else can be inferred when we see *described observation*?

Sense → Likelihood

- How likely is it that when *described observation*, it indicates that *described structure* is going on?
- How likely is it that when *describe structure* is happening, if we see *described observation*?

Action → Event

- As a result of *described action*, what might happen?
- Besides *other described resulting event, do you think that anything else might happen as a result of described action?*

Dependence between events→ Mutually-dependent event

- Does *described event dependence* become more or less likely in tandem with any other events?
- Do any other events cause *described event dependence* to become more or less likely?

Dependence between events→ Independent event

- What events does *described dependent events depends* depend on?
- Are there any other events which *described dependent events* depends?
- Which event cause *described dependent events* depends?

Dependence between events → Dependent event

- Are there any events that are made more or less likely by those occurring *described dependent events*?

Dependence between events→ Likelihood

- How likely is it that if *antecedent events in participant's words* occurs, that *consequent events in stakeholder's words* will happen.

Stakeholder → Anticipation

- And as a result of them experiencing that, do you think they will do anything in response?
- Are there any other possibilities for how they might act in response?
- What else do you think that *participant description of stakeholder* might do?
- Is it possible that *described stakeholder* could do something else in addition to *described anticipation*?

Anticipation (all kinds) → Stakeholder

- Who do you think will do this?
- Who might act in this situation?
- Is there anyone else who might do this?
- Is there anyone else who might act in this situation?

Anticipation (all kinds) → Dependence between anticipations

- Will this action be more or less likely if other actions are taken?
- Are there other actions that either increase or decrease the likelihood of this action being taken?

Anticipation (all kinds) → Action

- What do you think they will do?
- What will *described stakeholder* do?
- Is there anything else they might do?
- Is there anything else that *described stakeholder* might do?

Anticipation (all kinds) → Structure

- Under what conditions will they take this action?
- What conditions have to true for *direct quote naming stakeholder* to do *direct quote describing action*?

Dependence between anticipations → Independent anticipation

- Are there any actions that other stakeholders might take that would make *described action of other stakeholder* more or less likely?

Dependence between anticipations → Mutually-dependent anticipation

- What actions depend on this action to be taken or not to be taken?
- What actions do you anticipate not being taken if this action is undertaken?
- What actions will be taken if this action is undertaken?

Dependence between anticipations→ Dependent anticipation

- Are the actions of any other stakeholders dependent on this action being undertaken?

All items → Deference

- Is there anyone who knows about *described item more than you do, or alternatively, is there anyone that we should not listen to?*

Deference → Stakeholder

- Who is that?
- Who are they?

Deference → Likelihood

- How likely are they to understand this better than you?
- How likely are they to understand this worse than you?

Let us take a look at two of these questions to understand what they establish.

"When *described event* happens, besides causing *described result*, what else might occur as a result?"

This question is establishing the objective understanding of the stakeholder, and in particular their causal connectivity between events. This question provides the opportunity to develop how events could be more complicated than a simple story of cause an effect.

"When *described event* happens, who benefits/suffers as result?"

This question is establishing the subjective perceptions of the stakeholder. This question provides the opportunity for new stakeholders to be discovered.

Follow-up Interviews

After the initial interview, we might be less interested in discovering what stakeholders think and more interested in discovering how their understand reacts with findings from other stakeholders. At this point we can undertake a follow-up interview, or engage in follow-up interviewing in the initial interview session.

Of course, the kind of following-up that is appropriate depends upon one's methodological objectives. However, using this particular scenario elicitation and modeling practice allows one to engage concurrently in three research objectives that are traditionally understood to be mutually exclusive: understanding, reduction, and intervention [Braa and Vidgen, 1999]. Through these interviews, we come to understand the mental model of participants. This approach also allows the construction of probability models that afford a precise interpretation of risk. The intervention component of the model is the most subtle. By allowing participants to go on public record, it allows them to correct any view previously held about them incorrectly. Yet, at the same time to declare such a model is to make a speech act, which makes a commitment by which to judge their later action. This also allows those with oppositional viewpoints to make directed criticism about specific promises. At the same time, this forces opponents to articulate more specific accusations and suspicions. By sharing these models, participants are socializing their values and perceptions, which can help them constitute a broader strategy [Jones, 2007].

So, how does follow-up interviewing work in this particular interviewing strategy? First of all, it is useful to follow up if the interview reveals incompletenesses or contradictions between the participants. These incompletenesses and contradictions include:

- Another stakeholder successfully connects the models of a stakeholder, connecting different losses to the same overall situation.
- Another stakeholder expresses a major concern about an action the stakeholder

may be involved with that the stakeholder did not know.

- A trusted stakeholder belies the mental model of another stakeholder.
- A distrusted stakeholder demonstrates a shared understanding with the distrusting stakeholder about a point believed under contention.

Once one has discovered this kind of contradiction, one can ask such questions as "What about *stakeholder described by other participant*, would they experience *impact described by other participant* in this situation? What do you think they might do as a result?"

Of course, it is only appropriate to engage in these kind of follow-up interviews if the viewpoints of the participants have been completely explored, as best as the interviewer can be determined, and the point at which follow-up interviewing begins should be clearly denoted. The reason for this is to preserve the exchangeability constraint so that we can assess the what we have discovered honestly.

Interviewing Software Design

As the interviewing process is made substantially easier through software, potential users should understand the underlying design so that they can build and adapt versions for their own purposes. This design uses the classic Model/View/Controller implementation pattern which separates the data model (model) from the presentation of the data (view) from the logic needed to retrieve and process the data to make it suitable for presentation (controller). Let us look at each of these in turn.

The initial data model is simply a list for each model element which specifies what other elements can be connected to it via questioning, such as from impact to stakeholder. This specification also admits inheritance relationships, so that we can say "if x connects to y, z also connects to y", which simplifies matters a great deal when x has many connections. For example, all preconditions are structures, so since a structure can connect to an impact, a precondition can also connect to an impact.

As the interview is undertaken, another data model is populated, corresponding to the tree of interview questions and answers. Each node within the tree contains a question number and answer text, and each link between nodes represents a question referencing the parent node. Each node contains the text of the interviewees answer, as well as a check box to indicate that there were no further answers to that question. The tree starts with a root representing notes gathered before any question has been asked.

Having established the model, let us turn to the view. The overall interview data is viewed as a tree, whether the current pending answer is the node selected in that tree. The contents of the selected node are shown, and are presented as a text-box to write the answer, buttons for each kind of question that remains to be asked about it, and a check box if there is no answer. Also provided are open and save menu options for opening a previous model and saving the current model, respectively.

Given this model and view, we can talk about the controller. When the user clicks to ask a new question, a new node in the tree is created as the child of the node currently in focus, and then focus is transferred to that node. When the no-answer checkbox is clicked, the parent of the active node is checked to see if it has any possible further questions. If so, that becomes the active node, and if not its parent is checked in a similar way, recursively. If the save menu item is clicked, the tree is converted into a text file that summarizes each node of the tree, while open returns files of the same format.

Interviewing and Bias Avoidance

Given that we have described an interviewing process, it is worthwhile to see how this approach addresses our core concern, which is to use the psychology of how individuals understand counterfactual knowledge to address the biases of political expertise. First of all, by grounding ourselves strictly in the words of the participant, we avoid introducing support-theoretic biases. Next, by eliciting paths of events instead of asking for predictions, including the ways that they intercede and cancel, we help mitigate base-rate neglect. Further, by always asking for more, from the perspective of all of the elements involved, we avoid availability biases that might result from purely open-ended lines of questioning. Finally, by separating facts from concerns, we can gain an understanding of where individuals actually disagree. We will later see how the paths found through this process help mitigate risk governance deficits.

A Synthesis of Design Methods and Risk Governance Needs

Now that we have constructed a methodology for representation, simulation, discovery, and elicitation scheme, what can we say about it? First, we will both synthesize the work done so far, and analyze it carefully to show that we have met Tetlock's criteria, at least theoretically. Next, we will look at the project objectives, and show that these were met as well. Then, we will give a brief example of how to begin such a project by looking at some of the perspectives involved in the risk governance of sequestration-based carbon mitigation technologies. Following this, we will provide a tiny example of how this analysis works in practice. Finally, we will look at the future work, and see what potential directions this work can take next.

Analysis and Synthesis

In this section, we take on the criteria that Tetlock provided for debiasing methodologies such as the one we have developed. To review, the criteria is given here:

> *Promoters of "debiasing" schemes should shoulder a heavy burden of proof. Would-be buyers should insist that schemes that purportedly improve "how they think" be grounded in solid assumptions about (a) the workings of the human mind and -in particular- how people go about translating vague*

> *hunches about causality into the precise probabilistic claims measured here; (b) the workings of the external environment and -in particular- the likely impact of proposed correctives on the mistakes that people most commonly make in coping with frequently recurring challenges.*
> From *Expert Political Judgment* by Philip Tetlock [Tetlock, 2005b]

First, we will take on criteria (b), and show that these methods help cope with frequently recurring challenges by showing how risk governance deficits lead to risks, and how the methods we describe intercede. Second, we will take on (a), by showing that the causal paths of our model can be converted into precise probability scores, and further offer another analytical tool, which is a measure of the expert's assessment of the contingency of predicted events. In the course of this discussion, we will demonstrate a synthetic view of how these methods might operate in the context of risk governance.

Modeling Risk Governance Deficits

The Risk Governance Council supplies at total of twenty three risk governance deficits, ten of which (labeled A1 to A10) are devoted to assessing and understanding risks, while the remaining thirteen (labeled B1 to B13) are devoted to managing risks. As mentioned previously, this paper takes these deficits as hard-won empirical observations from risk governance failures, and therefore they serve as appropriate guidelines for shaping risk governance approaches.

Using these deficits, we can attempt to respond to one of the two criteria that Tetlock's *Expert Political Judgment* sets for us earlier, namely *"Would-be buyers should insist that schemes that purportedly improve 'how they think' be grounded in solid assumptions about . . . the workings of the external environment and -in particular- the likely impact of proposed correctives on the mistakes that people most commonly make in coping with frequently recurring challenges."* [Tetlock, 2005b]. If we limit ourselves to the risk governance context and we accept this list as the mistakes faced within risk governance, then should we be able to show that measures could help with these deficits, we can consider this obligation met.

How is it that these deficits can be linked to the formalization we have described so far? One way to understand this is if each deficit represents either a path to an impact, or a missing link in an intervention preventing a path to impact. If we can then represent a path disrupting the impact by building an intervention, then we can talk about procedures aimed to elicit or induce paths of intervention, while simultaneously not compromising other such paths. This approach is very familiar to risk management audiences, finding one of many other manifestations in fault-trees (see [Leveson, 1995] for a fine survey).

Let us now describe a way to construct these kinds of paths[17]. We can describe this path as a repeated application of relations. For example, suppose that from the current conditions (s), we correctly make an observation of the current conditions ($\theta(s) \mapsto o$), which leads us to take a different anticipation of how to act ($an(o) \mapsto a$), which leads to a sequences of events occurring over some period of time ($e^*(s,a) \mapsto (s_2, t)$), which leads to a different set of impacts for some stakeholder $R(s_2, sk) \mapsto \Re^c$, as desired. We can then talk about that entire string of relationships in a single expression, as in $s; \theta(s) \mapsto o; an(o) \mapsto a;\ e^*(a,s) \mapsto (s_2, t); r(s_2, sk) \mapsto \Re^c$. We can now represent risk governance deficits as disruptions in these intervention paths, for which we will use a 'hat' notation (such that a failure to interpret an observation as a symptom of a potential risk would be $\hat{\theta}$). We could then represent impacts caused by such a misinterpretation as $s; \hat{\theta(s)} \mapsto o; an(o) \mapsto a; e(s,a) \mapsto s_2;\ r(s_2, sk) \mapsto \Re^c$. By convention, we shall take terms separated by ; and , to be effectively sequential and concurrent, respectively.

Given this formalization, let us examine the twenty-three risk governance deficits and see which of them are easily formalized, to see if the basic framework theory presented earlier interprets them sensibly.

A1: *The failure to detect early warnings of risk because of erroneous signals, misinter-*

[17] If the notation of these paths needs more explanation, it may be handy to consult Appendix E.

pretation of information, or simply not enough information being gathered

This deficit could be described as a deficiency in knowing of potential observations, mapping those observations to structures correctly, or in practically capturing those observations. There are number of ways this deficiency could manifest itself. Perhaps we did not capture knowledge of the observation itself (we were missing an $o \in O$). Another way to make a mistake here is to misinterpret what the observation entails: say we did not associate that observation with the state that was actually implied (we missed a structure/observation mapping in Θ), or we misspecified the probabilistic elements of the mapping, leading us to misinterpret the information we had (the distributions in Θ are wrong). Finally, even if the general knowledge is correct, the specific observations that were reported could be erroneous, either falsely observed or omitted (O_{sk} is misspecified).

Given this, a path to risk caused by this deficit could be formalized as $s; \hat{\theta}(\hat{s}) \mapsto \hat{o}; an(o) \mapsto a; e(a) \mapsto s_2; r(s) \mapsto \Re^c$.

A knowledge discovery procedure aimed to eliminate this deficit should inquire, in all relevant states, what one would observe in those states, what other states one might also make the same observations, what factors would make the states more likely relative to each other, given those observations; and which of these observations are currently observable. The interview procedure we have described meets this requirement.

A2: *The lack of adequate factual knowledge for robust risk assessment because of existing gaps in scientific knowledge or failure to either source existing information or appreciate its associated uncertainty*

We can characterize this deficit as failing to understand correctly how events transition between structures. As before, this failing manifests itself in multiple ways. This could be due to the omission of events entirely (missing an $e \in E$), misunderstanding the causes that events turn out differently (misspecifying the se in the input of an event, e), neglecting or misunderstanding the consequences of an event (misspecifying the

se in the output of an event, e), mischaracterizing the likelihood of events causing between structures in particular ways (misspecifying the event distribution, $P(E)$, or dependencies between events), or mischaracterizing the dynamics of structures.

The path to harm described by this deficit is: $s, \theta(s) \mapsto o; \theta^{-1}(o) \mapsto s; \hat{e^*}(\hat{s}) \mapsto \hat{s_2}; r(s_2) \mapsto \Re^c$.

A knowledge discovery procedure aimed to eliminate this deficit should inquire, in all relevant states, what events could occur those states, what other states one might also make the same observations, what factors would make the states more likely relative to each other, given those observations; and which of these observations are currently observable. The interview procedure we have described meets this requirement.

A3 : *The omission of knowledge related to stakeholder risk perceptions and concerns*

The most basic omission of stakeholder concerns is to neglect an entire set of criteria (missing a $c \in C$). It is also possible to misspecify the magnitude of one or many of those concerns (misspecifying a $\Re_c$ in $\Re^C$). In addition, it is possible to make mistakes about the stakeholders affected (misspecifying $sk \in Sk$), as well as the conditions under which various impacts occur (misspecifying $s_2 \in S$).

Such misspecification could cause this path to risk: $s, \theta(s) \mapsto o; an(o) \mapsto a; e^*(a, s) \mapsto s_2; r(\hat{s_2}, \hat{sk}) \mapsto \hat{\Re}^{\hat{c}}$.

A knowledge discovery procedure aimed at eliminating this deficit would inquire, in all relevant states, what impacts could occur, which stakeholders would be affected, how would those stakeholders be affected, and to what degree those stakeholders would be affected or how the effect of the impact should be quantified. The interview procedure we have described meets this requirement.

A4: *The failure to consult the relevant stakeholders, as their involvement can improve the information input and the legitimacy of the risk assessment process (provided that interests and bias are carefully managed)*

Failing to consult stakeholders can be understood as a failure to take an action involving a stakeholder $(e(a, sk) \in E : (a, sk) \in \mathcal{P}(Se))$, leading to an event undertaken incorrectly. Given this, a path to risk described by this deficit is $s; \theta(s) \mapsto o; an(o) \mapsto \hat{a}; \hat{e}^*(\hat{a}, \hat{sk}) \mapsto \hat{s_2}; e^*(a, s_2) \mapsto s_3; r(s_3, sk_2) \mapsto \Re^c$.

A knowledge discovery procedure aimed at eliminating this deficit would need to include stakeholders that incorporate both the diversity of relevant knowledge and the breadth of relevant value experience. For example, in site-selection for some new infrastructure, relevant knowledge could be concerned with both the general scientific understanding of the benefits and risks imposed by that kind of infrastructural installation as well as specific local knowledge about the site. By adding the requirement that interviews must be taken with all discovered stakeholder categories, the interview procedure we have described meets this requirement.

Value experience is also vital to producing legitimate risk assessments. knowledge discovery procedures need to consult *both* the stakeholders impacted by risk governance decisions, as well as those with relevant prior interactions, experience, and expertise with the kinds of harms experienced. Both is highlighted, as it reflects a particular kind of ambiguity faced by the risk domain. Consider legislation requiring wearing motorcycle helmets. A motorcyclist with a libertarian bent may insist on their right not to wear such a helmet and insist that they would forgo any treatment in the event they are injured. A physician with experience treating motorcycle injuries may equally insist that it this is inappropriate, as those employed in medicine work under a professional ethic where care is paramount and an institutional context where treatment is not withheld. Both of these parties have relevant, but not total, knowledge toward arbitrating this claim. An individual who experienced such an accident, and who either changed their stance or who persevered with their view at personal cost, would bring even stronger relevant evidence, although it would still not be definitive.

A5: *The failure to properly evaluate a risk as being acceptable or unacceptable to society*

Acceptability is an extremely curious idea. Identical risks, arbitrated through identical measures, will have different levels of acceptability to different individuals. Therefore, to violate the acceptability of a risk is to cause a second-order harm, namely to offend cultural sensibilities. Conversely, if a risk is acceptable, but treated otherwise, then mitigation activities will be seen as wasteful, and the cost of which will be seen as a loss by stakeholders. Even though offenses to acceptability can lead to other harms, such as the social amplification of risk and the loss of institutional credibility and support, transgressions of acceptability are harms in themselves.

Given this, a path to a preventable loss can be described by the following path: $s; \theta(s) \mapsto o; an(o) \mapsto a; e^*(a, sk) \mapsto s_2; r(s_2, sk) \mapsto \Re^{c_1}; \Re^{c_1} \mapsto s_3; s_3 \mapsto r(s_3, sk_2); \hat{r}(s_3, s\hat{k}_2) \mapsto \hat{\Re}^{\hat{c_2}}$. As mentioned before, the single most interesting about this kind of path to risk is that there may be no deficit between the phenomena all the way to and including the primary risk itself. An important factor that this path leaves out is the other factors that determine risk acceptability. It is insufficient for a harm merely to be conferred for that harm to be unacceptable, but often must be accompanied by an inability of those suffering the harm to have either known of it or have acted to prevent it. Symmetrically, knowledge that a stakeholder was capable of preventing the harm, but did not, can prevent the risk being deemed unacceptable. It is also possible for those experiencing the acceptability harm to be mistaken about these additional acceptability factors. In any case, s_3 should be interpreted broadly to potentially incorporate these potential understandings and misapprehensions.

To mitigate this risk governance deficits, knowledge discovery procedures should not stop at the harm, but then ask who may come to observe the harm, as well as the actions of those experiencing harm and their supporters. It is essential to remember that the implications of a harm never stop with the harm itself. The interview procedure we have described meets this requirement.

A6: *The misrepresentation of information about risk, whereby biased, selective or in-*

complete knowledge is used during, or communicated after, risk assessment, either with or without intention

It is reasonable to construe the communication of information as an action that causes the event of one state of knowledge being transformed into other. For this reason, it is appropriate to model the selection and dissemination of knowledge as communicative actions. Given this, we can treat misrepresentations as flaws in the actions of those selecting or disseminating the information (mistakes in choosing the actions $a \in A$ of communicative events $e \in E$, leading to errors in the state of knowledge of $s_2 \in S$). The corresponding risk pathway is $s; \theta(s) \mapsto o; an(o) \mapsto a; \hat{e}^*(\hat{a}, s) \mapsto (\hat{s_2}, \hat{t}); an(sk, \hat{s_2}) \mapsto a_2; e^* a_2, s_2 \mapsto s_3; r(s_3, sk_3) \mapsto \Re^c$

In order to avoid this deficit, a knowledge discovery procedure should aim to eliminate two different kinds of errors. First, to be assured that one discovers and selects information correctly, or that is to say, that one communicates information to oneself in an unbiased way. This requires explicitly asking about how information was discovered, including the prior conceptions, new conditions, and actions which lead to it. Second, to be assured that communicative actions are having the intended effect on those receiving the information. Such a method may be more effective when it attempts to make sense of the domain alongside those it is communicating to, rather than attempting to see merely if a transmitted message has been retained [Dervin, 2001]. By remaining domain neutral, the interview procedure we have described meets this requirement.

A7: *A failure to understand how the components of a complex system interact or how the system behaves as a whole, thus a failure to assess the multiple dimensions of a risk and its potential consequences*

It is entirely possible to understand the local dynamics of each event perfectly, including its preconditions, distribution of likelihood, distribution of duration, and dependencies upon other events, yet still fail to understand the behavior of a complex system. This is because this knowledge does not explicitly construct the pathways between events, nor does it posit which phenomena may come to interact when produced through independent pathways. For this reason, we should note that although there is little notational difference between misunderstanding the behavior of given events in a particular timeslice ($\hat{e}^*$) and misunderstanding the interactions between events across a timeslice ($\hat{e^*}$), the difference between these understandings is profound. We could equally well describe this deficit as misunderstanding the dynamics of sequences of dependencies (d^*). With these caveats, we describe this risk pathway as $s; \theta(s) \mapsto o; an(o) \mapsto a; \hat{e^*}(a, s) \mapsto (s_2, t); r(s_2, sk) \mapsto \Re^c$.

It is in the interaction of complex systems that computational formalisms show their strength, due to the ability to search and simulate combinatorial possibilities. Although a well-known catalog of systemic interactions can be recognized through more straightforward modeling activities, it is progressively more difficult as systems grow larger and more dynamic in their interactivity. Elicitation alone is likely insufficient in this case, although it is always important to ask about the dependences ($d \in D_E$) between events. However, by providing a simulation procedure, our method meets this requirement.

A8: *A failure to recognize fast or fundamental changes to a system, which can cause new risks to emerge or old ones to change*

In some ways, this deficit is a curious one, in that it conflates two kinds of system changes (*fast and fundamental*) that are apparently very different. However, both of these are failures to observe changes at a timescale appropriate to deal with them after initial observations have been made. These deficits could come from failing to recognize the possibility of a new condition ($s_2 \in S$), failing to recognize a new temporal regime for the pace of events ($t \in T$), failing to observe a new condition ($o \in O$), or misinterpreting that observation as not providing the correct warning of new conditions ($\theta \in Theta$). These correspond to the following path to risk: $s; \theta(s) \mapsto o; an(o) \mapsto a; \hat{e^*}(a, s) \mapsto (\hat{s_2}, \hat{t}); \hat{\theta}(\hat{s_2}) \mapsto \hat{o}; an(o) \mapsto a_2; \hat{e^*_2}(a_2, s_2) \mapsto (s_3, t_2); r(s_3, sk) \mapsto \Re^c$

To mitigate against this deficit, knowledge discovery procedures should, in addition to asking about all possible event dynamics, continue to ask about potential observations that may reveal those underlying changes. The interview procedure we have described meets this requirement.

A9: *The inappropriate use of formal models as a way to create and understand knowledge about complex systems (over- and under-reliance on models can be equally problematic)*

The discussion in the section *"Why Technical Modeling?"* also applies to discussions of this deficit. In the middle of formalizations it is worthwhile to take a brief pause to remind the reader that the purpose of formalization is always to generate new insights about governing risk, and the use of formalization is to efficiently generate insights where more straightforward reasoning becomes bogged down in complexity. As it stands, to use models inappropriately is to develop and act upon a state of knowledge (s_2 in S) inappropriately. In this way, this deficit is very similar to biases in selecting and communicating information (A6), and therefore shares the same risk pathway: $s;\theta(s) \mapsto o;an(o) \mapsto a;\hat{e}^*(\hat{a},s) \mapsto (\hat{s_2},\hat{t});an(sk,\hat{s_2}) \rightarrow a_2;e^*a_2,s_2 \mapsto s_3;r(s_3,sk_3) \mapsto \Re^c$. However, by providing a model for the discovery process generated by the model, this method provides a way to critique the method's appropriateness, and thus helps mitigate this deficit.

A10: *A failure to overcome cognitive barriers to imagining that events outside expected paradigms are possible*

While it would be possible to characterize missing any of the terms in this model as possibly being a sign of cognitive barriers, there are definitely some effects worth noting that correspond well to the model itself. In particular, there are many opportunity to look for signs of integrative complexity, or lack thereof. First of all, if the mental model of an individual is dominated by one-to-one relationships when many-to-many are allowed. Some of the possible many-to-many relationships include multiple results of a given event, multiple causes of an observation, multiple observations corresponding to a given phenomena, multiple possibilities of anticipated response, and multiple impacts of a given state-of-affairs (to multiple stakeholders, along multiple criteria). In the same vein, another sign of integrative complexity is expressing uncertainty or giving probabilistic judgments instead of making deterministic remarks.

In any case, a knowledge discovery procedure can guard against this by always asking for further possibilities, until the stakeholder can think of no further responses. In modeling, setting high priors on the number of possibilities can against closing models too soon. Furthermore, an explicitly open modeling paradigm that makes no assumptions as to its own finality, as well as the operational support to continually revisit and revise assumptions, allows for modeling to become a tool against cognitive barriers instead of a static asset who's viability depends upon maintaining them.

However, we can make even stronger assurances against outside-of-paradigm events, by engaging in the analysis of the discovery activity itself, and attempt to assess its current progress in an online fashion. To the degree that these tools are applied and maintained continuously, this method can assist with such assurances.

B1: *A failure to respond adequately to early warnings of risk, which could mean either under or over-reacting to warnings*

It is entirely possible to understand a warning sign correctly, yet react incorrectly to it, yielding a chain of events that undermines risk mitigation activities. For this reason, there may be a misspecification of the action that should be undertaken ($a \in a$). The resulting risk pathway is $s; \theta(s) \mapsto o; an(o) \mapsto \hat{a}; e^*(a, s) \mapsto (s_2, t); r(s_2, sk) \mapsto \Re^c$

In order to guard against this risk, a knowledge discovery strategy should ask about all available actions to early warning observations, and follow up on the potential consequences of those actions. The interview procedure we have described meets this requirement.

B2: *A failure to design effective risk management strategies. Such failure may result* *from objectives, tools, or implementation plans being ill-defined or absent*

What does a failure to design risk management strategies mean? In this context, design needs to be understood as a response to observable factors. It is incoherent to say "we need to design a risk management strategy that causes unknown harms, in an unknown way, according to an unknown mechanism", as that offers no clue of what the identity of the problem may be. However, as observations are made, that may incite novel discoveries, combinations, and concepts. Therefore, we can say that a failure in design is to fail to take action to generate the necessary planning and knowledge from more poorly understood knowledge, such that when observations of a developing situation occurs, we have made measures to be prepared. Thus, when we have a chance to react to observable factors, and fail to design appropriately ($\hat{a}$), this leads to us fail to gather those designs ($\hat{e}$), leading to a risk pathway of $s; \theta(s) \mapsto o$; $\hat{e}^*(o, \hat{a}) \mapsto (s_2, t)$; s_2; $\theta_2(s) \mapsto o_2$; $an(o) \mapsto a$; $e^*(a, s_2) \mapsto s_3$; $r(s_3, sk_3) \mapsto \Re^c$.

Knowledge discovery strategies play two roles here. The first is to ask how observations about situations are being converted into plans handling their risk. However, a second role of a knowledge discovery strategy is to act as a design strategy. That is to say that knowledge discovery procedures, among which our procedure is a member, are design methodologies themselves, designed to address this deficit.

B3: *A failure to consider all reasonable, available options before deciding how to proceed*

Although failing to consider all possible actions seems to be a straightforward failure in building anticipations and their resulting actions correctly, there are notable variations. These include not being aware of the existence of potential actions (missing $a \in A$), not recognizing known actions as being applicable to the present conditions (misspecifying the anticipation, an, by not linking the action as being applicable given the current observations), not recognizing the saliency of actions to changing events (misspecifying the actions in the precondition of events, as in $e^*(\hat{a}, s)$), and not recognizing the possibility of not acting at all.

Given this, a risk pathway corresponding to this deficit is $s; \theta(s) \mapsto o$; $\hat{an}(o) \mapsto \hat{a}$; $e^*(\hat{a}, s) \mapsto (s_2, t)$; $r(s_2, sk) \mapsto \Re^c$

A knowledge discovery procedure that would guard against such a risk would ask about which actions are available to all stakeholders under all conditions, and would ask again after some number of actions had been mentioned, until the stakeholder providing the information could think of nothing further. Our method does exactly this.

B4: *Not conducting appropriate to assess the costs and benefits (efficiency) of various options and how these are distributed (equity)*

Most risk governance decisions will require trade-offs. For the cost of some precaution, we can enjoy freedom from a particular class of risk. Sometimes these trade-offs are very straightforward: for a given expense, a given stakeholder decided that they are willing to pay to mitigate a certain risk, but then discover that they could instead use those funds to eliminate a more dangerous risk. However, in the aggregate, it can be the case that different individuals benefit from different risk reductions. In this case, we may be treating the overall benefit as a comparison between statistical aggregate, with corrections assigned for the distribution of who endures various risks and expenses. In both matters of efficiency and equity, we are judging one class of risks and versus another, and therefore the deficit does not necessarily arise from the failure to assess or respond to either of the losses being traded between, but to judge that trade-off correctly. We can say that between the losses between two stakeholders, there has been a significant imbalance (an imbalance more significant than some negligible quantity within a margin for error, ϵ) in the judgment of a stakeholder ($\Re^c_{sk_1} \hat{<}_{\epsilon, sk_3} \Re^c_{sk_2}$). The corresponding risk pathway is described as: $s; \theta(s) \mapsto o; an(o) \mapsto a; e^*(a, s) \mapsto (s_2, t); r(s_2, sk) \mapsto \Re^c_1, r(s_2, sk_2) \mapsto \Re^c_2, \Re^c_1 \hat{<}_{\epsilon, sk_3} \Re^c_2$.

Knowledge discovery strategies that are designed to address this deficit include always asking about the specific magnitude and likelihood of particular risks, including follow-on costs and asking about the resources needed for the various mitigation strategies proposed, as well as their risks. Analysis strategies for uncovering these deficits include examining the risk portfolio held by each stakeholder category, as well as their contributions to risk-eliminating resources. It should also be appropriate to attempt to characterize the norms of each community in equity questions. The methods described here undertake all these measures.

B5: *A failure to implement risk management strategies or policies and to enforce them*

The failure to implement or enforce strategies is the analogue of risk deficit B2, the failure to design risk management strategies. Here, the actions after the first hint of the risk are just fine, but when an observation that they need to be undertaken arrives, the action that was designated is either not taken or taken poorly, leading to events failing to manage risk. Such a risk pathway could be described as: $s; \theta(s) \mapsto o; e^*(o,a) \mapsto (s_2,t); s_2; \theta_2(s) \mapsto o_2;\ an(o) \mapsto \hat{a}; \hat{e}^*(\hat{a}, s_2) \mapsto s_3; r(s_3, sk_3) \mapsto \Re^c$. Outside of ordinary human folly, this could also result from a miscommunication or misunderstanding between those designing and implementing these strategies.

A knowledge discovery strategy designed to mitigate this risk is to not only engage those planning the risk response, but also to those responsible for implementing and enforcing these measures. It also cannot hurt to check about the resources needed for particular strategies, and the guarantees of their availability. The methods described here do not directly address this particular problem, although they can aide risk policy implements to think about potential contingencies.

B6: *A failure to anticipate the consequences, particularly negative side effects, of a risk management decision, and to adequately monitor and react to the outcomes*

A risk management strategy, well-designed and competently executed, can lead to failures if the unintended effects are not mitigated. In this risk pathway, we describe the results of actions as producing an unexpected state-of-affairs $(\hat{s_2})$ leading to loss: $s, \theta(s) \mapsto o, an(o) \mapsto a, e^*(a,s) \mapsto (\hat{s_2}, t), r(s_2, \hat{sk_2}) \mapsto \Re^{\hat{c_2}}$

A knowledge discovery strategy designed to eliminate this deficit would ask about the various consequences of mitigation actions, for all stakeholders. Further, it would ask if there are stakeholders tasked with monitoring the results in the event the effects could be unknown. The elicitation method does ask about the extended outcomes of particular paths, so it may offer a slight contribution. Of course, ongoing storage and simulation of possibilities may assist monitoring in being mindful of contingencies. Overall, however, the effective use of these tools for monitoring purposes is entirely contingent on their discretionary use by those with operational responsibilities.

B7: *An inability to reconcile the time-frame of the risk issue (which may have far-off consequences and require a long-term perspective) with decision-making pressures and incentives (which may prioritize visible, short-term results or costly reductions)*

This deficit is one of the core issues of distributed risk problems. One way to conceive of this is deficit is as a variation on B4, or a mistake in analyzing the effectiveness or equity of an action. Here, risk mitigation actions lead to two different sets of benefits and losses, which would again be judged to have a substantial disparity ($\Re^c_1 \hat{<}_\epsilon \Re^c_2$), but this time, in addition, the time between these effects is substantial ($t_1 >_\epsilon t_2$). As an additional complication, this disparity might not be observable to any stakeholder, given the time-difference between them, and might only be assessed in the abstract. Given this, we can describe the risk pathway as: $s; \theta(s) \mapsto o; an(o) \mapsto a; e^*(a,s) \mapsto (s_2, t_2) \wedge (s_3, t_3); r(s_2, sk) \mapsto \Re^c_1, r(s_3, sk_2) \mapsto \Re^c_2, \Re^c_1 \hat{<}_\epsilon \Re^c_2, t_1 > t_2$.

For a knowledge discovery strategy, the first line of defense is to establish the duration of the effects of events, being sure to capture those that are a long distance away. Additionally, it is important to account for scientific knowledge that, although not giving us direct insight, allows some demographic projections about the stakeholder communities likely to be affected. Of similar importance is eliciting information from institutional stakeholders, who may be able to stand in for ongoing populations of similar stakeholders. Economic strategies to understanding the value of future losses also provide helpful tools, through the notion of option value [Posner, 2004] [Sunstein, 2007], or the right to a particular resource, for a fixed price, at a later time, such that potential future losses can be described as losses of current assets (namely the futures contracts in question). The methods in this paper, by using a formulation that allows for the implementation of neuro-dynamic programming methods, allows for simulations to effectively propagate probabilistic risk, and thus can be modified to address this challenge.

B8: *A failure to adequately balance transparency and confidentiality during the decision-making process, which can have implications for stakeholder trust or for security*

The boundary between confidentiality and transparency is a tricky wrinkle in the already difficult challenge of managing communication actions. In order to understand the dynamics of this, let us work our way through the risk pathway: $s; e^*(sk_1, a_1) \mapsto s_2; \hat{e}^*(sk_2, \hat{a_2}, s_2) \mapsto \hat{s_3}; e^* sk_3, a_2, s_3 \mapsto s_4; r(s_4, sk_4) \mapsto \Re^c$

Here, a stakeholder (sk_1), either deliberately or inadvertently, discloses (a_1) some information, leading to a state where those attempting to govern risk know some additional aspect of information (s_2). This information is then either inappropriately disclosed or inappropriately withheld ($\hat{a_2}$) by the stakeholder to whom the information was disclosed (sk_2). This disclosure leads to a state of knowledge where this information is either inappropriately known or unknown ($\hat{s_3}$). As a result of this state, a different stakeholder (sk_3) act differently according to that knowledge, leading to a harm dealt to a stakeholder (sk_4), who may be any of the previous stakeholders.

In order for this to be navigated directly, those using knowledge discovery procedures should be very careful to make clear exactly how the knowledge will be used. Another concern is accidentally revealing knowledge in the knowledge discovery processes. This is a significant advantage for open-ended interviewing, grounded theory, or non-parametric methodologies aiming at participant-specific discovery and exchangability (such as those described here), as the formation of questions includes no domain-specific knowledge, and thus no chance of unintended cross-communication.

B9: *A lack of adequate organizational capacity (assets, skills, and capabilities) and/or of a suitable culture (one that recognizes the value of risk management) for ensuring managerial effectiveness when dealing with risks*

One way to think of managerial effectiveness is as the skill of cultivating stakeholders; to shepherd them into mutually suitable knowledge and interests, and to see that this knowledge and interest is maintained and further cultivated. So, a failure to develop organizational capacity is a failure to take the actions that lead from one set of stakeholders becoming another, or remaining competent as another ($\hat{e}^*(\hat{a}, \hat{sk}_1) \mapsto (\hat{sk}_2)$), leading to an overall risk pathway of $s; \theta(s) \mapsto o; an(o) \mapsto a; \hat{e}^*(\hat{a}, \hat{sk}_1) \mapsto (\hat{sk}_2); e^*(sk_2, a, s) \mapsto (s_2); r(s_2, sk) \mapsto \Re^c$.

A knowledge discovery strategy aimed at discovering this deficit is always to inquire who will be undertaking risk mitigation, either the design, implementation, enforcement, or monitoring. This is another reason why it's particularly important to attempt to interview relevant groups of stakeholders, as they can be found not to yet exist. This method addresses this deficit by incorporating risk mitigation actions, events, and stakeholders within these models.

B10: *A failure of the multiple departments or organizations responsible for a risk's management to act individually but cohesively, or of one entity to deal with several risks*

Another way to consider this deficit is as a failure in coordination. This means that at least one stakeholder fails to act in a way that takes another action into account, or $\hat{e}^*(\hat{a_1},\ \hat{sk}_1, \hat{a_2},\ sk_2, s)$, leading to a risk pathway of $s; \theta(s) \mapsto o$; $an(o) \mapsto a$; $\hat{e}^*(\hat{a_1}, \hat{sk}_1, \hat{a_2}, sk_2, s) \mapsto (s_2, t)$; $r(s_2, sk_3) \mapsto \Re^c$.

A knowledge discovery strategy aimed at understanding this deficit is to consult stakeholders engaging in risk management activities, to make sure that they are aware of the actions being undertaken throughout their organization, as well as in other organizations with related activities. In other words, an effective knowledge discovery practice may have to take multiple samples even within a single organization to come to an effective picture. This can be addressed in this method by incorporating risk mitigation activities into the overall risk model.

B11: *A failure to deal with the complex nature of commons problems, resulting in inappropriate or inadequate decisions to mitigate commons-related risks (e.g. risks to the atmosphere or oceans)*

Commons-related risks pose a special challenge to coordinating risk governance activities, due to the large number of stakeholders with an inescapable interest in, and claim to, a common resource. I've chosen to represent this risk as an unplanned sequence of events resulting from many populations of stakeholders taking many actions to a single state, or $\hat{e^*}(\hat{a^*}, s, \hat{sk^*})$. Given this formulation, a path to a potential harm is $s; \theta(s) \mapsto o; an(o) \mapsto a; \hat{e^*}(\hat{a^*}, s, \hat{sk^*}) \mapsto (s_2, t); r(s_2, sk) \mapsto \Re^c$.

Commons-related risks pose a special challenge to the scalability of risk governance methods, as everyone has a stake in the condition of the atmosphere. For this reason, the place to begin is likely at the largest institutional representative for a population of stakeholders, such as the responsible departments of governments. They could then direct one to specific groups with conflicting interests within their domain, as well as to known issues they are currently engaged in with their peers in other jurisdictions. Of equal importance is discovering those on top of emerging commons issues with little institutional representation, such as the international jurisdiction for climate geoengineering measures [Cascio, 2010]. These methods, by allowing for distributed elicitation, allow for model building to scale across commons.

B12: *A failure to resolve conflicts where different pathways to resolution may be required in consideration of the nature of the conflict and of different stakeholder interests and values*

In a certain way, this deficit does not have a risk pathway, in the model we have described, exactly because this deficit is mistake of one stakeholder in understanding the pathways as understood by another stakeholder, and in particular misunderstanding the misunderstandings of the other stakeholder. In some ways, more than any other, it is in this kind of deficit where elicitation processes matter. Examples of this deficit include mistakes regarding the subjective assessment of other stakeholders, particularly confusing objective validity with subjective assessment, which this protocol is designed to unravel.

However, when understood in a different way, this deficit does have a clear path to risk. Namely, this is a mistake in the kind of conflict resolution actions applied ($\hat{a}$) due to misunderstanding the mental model of the stakeholders in question ($\hat{sk}$), so we can write a risk pathway as $s; \theta(s) \mapsto o; an(o) \mapsto a; \hat{e}^*(\hat{a}, s, \hat{sk}) \mapsto (s_2, t); r(s_2, sk) \mapsto \Re^c$, even if to do so is to hide the mechanisms of this deficit. Given this, one way to understand the value of open-ended elicitation aimed at recovering the difference between objective and subjective understandings is to aide in the selection and design of the appropriate conflict resolution channels.

B13: *Insufficient flexibility or capacity to respond adequately to unexpected events because of bad planning, inflexible mindsets, and response structures, or an inability to think creatively and innovate when necessary*

The simple way to state such a deficit is as poor anticipation ($\hat{an}$) and action ($\hat{a}$) to an unplanned state-of-affairs (s_2), resulting from an unexpected sequence of events, as in the risk pathway $e^*(s) \mapsto s_2; \hat{an}(s_2) \mapsto \hat{a};\ e^*(a) \mapsto s_3; r(s_3, sk) \mapsto \Re^c$. However, this largely fails to capture the flavor of this problem. Flexibility and creativity are in some ways developed capacities. In this way, deficits in this area, to the degree that they can be mitigated at all, are developed through practices that build flexibility and creativity.

For this reason, it is the use of the results of elicitation and analysis that is important to mitigating this deficit. Stakeholders should, to the degree that privacy and effectiveness concerns allow, have a chance to be given alternatives discovered from others, outside of their current perspective, and come to terms with these results.

Overall, what does this exercise of mapping risk governance deficits into risk paths tell us? There are two conclusions. First, we see that the model presented here can usually capture the risk imposed the deficit, but at the cost of losing some of the semantics of the original deficit. This has the advantage of still being completely relevant to those stakeholders who are not involved with those particular aspects of the risk governance process, but loses some inductive power in putting relevant questions to those who are. Altogether, we can say that this domain-general framework of objective and subjective causal knowledge would strike an appropriate balance if it recovers those criteria when they are relevant to the stakeholder even without explicit mention, while discovering the potential of deficits even if the stakeholder does not think about their concerns in those terms.

Second, we can also say that many of the concerns expressed in these deficits are better resolved not through direct formalization, but through methodological heuristics in how to structure the process of elicitation. This is entirely appropriate, as the central question is how to structure a knowledge discovery and analysis process within the larger processes of risk governance.

Addressing the Deficits: A Summary

Given all this, it is useful to go back and see how the methods describe here addresses these risk governance deficits. This is the dual view of the above information: instead of describing, for each deficit, how these methods address them, here we are describing, for each method, which deficits they address. Given that we have already described this information in a different form, this section will aim for brevity, and will list only the most direct ways in which the deficits are addressed. personnel Interviewing: A1 (inquires into all possible observations given underlying conditions and their likelihoods), A2 (inquires into all possible events, results of those events, likelihoods of those results, and duration of those events), A3 (inquires into all stakeholders, their concerns, and the magnitudes of those concerns), A4 (discovers additional stakeholders), A5 (asks about follow-on “acceptability” harms), A8 (asks about observations revealing changes to system dynamics), B1 (asking about all available actions responding to observations that indicate early warnings), B2 (inquiring into current design activities for risk governance), B3 (asking about all available actions), B4 (inquiry into all costs, including mitigation costs, as well as stakeholder norms), B6 (inquires into the actions of all stakeholders including monitoring), B7 (asks with no fixed ending to the timespan inquired into), B8 (addresses each stakeholder separately with domain neutral questions)

Coding: A6 (avoids bias through domain neutrality and separation of objective/subjective concerns), A9 (guards against ambiguities due to insufficient formalization), B12 (separates objective and subjective concerns)

Inference: A4 (discovers the need to inquire into more stakeholder interests), A10 (inference of 'unknown unknowns'), B3 (inferring unmentioned actions), B10 (multiple samples within and across organizations)

Simulation: A7 (discovers unknown complex interactions by attempting to characterize patterns of interaction), B7 (simulates for the entire duration in which discounted risks remain)

Narration: A9 (guards against implausible codings and simulations by sanity checking simulation trajectories)

Re-interviewing: A13 (confronts fixed mental models with the results of other elicitation, allowing for reflection)

Overall: B2 (acting as a design method for building risk-governance strategies), B5 (discovering lack of implementation or enforcement), B9 (attempts to engage all stakeholder groups), B11 (engages every stakeholder independently, allowing for the process to scale)

Not addressed: B5 (method does not actually enforce or implement mitigation measures)

The Relation between Classical and Causal Bayesian Norms

Let us now look at the other criteria that Tetlock's *Expert Political Judgment* set for us earlier, namely *Would-be buyers should insist that schemes that purportedly improve "how they think" be grounded in solid assumptions about the workings of the human mind and -in particular- how people go about translating vague hunches about causality into the precise probabilistic claims measured here.* [Tetlock, 2005b].

This criteria is looser than it appears. We make no claim that this work corresponds directly to the 'the workings of the human mind'. The cognitive science we reference does not address the working of the human minds directly, but instead looks at computational models of human behavior. However, we can say that this science has generated 'solid assumptions about the working of the human mind' by virtue of replicating human behavior in controlled experiments. This research indicates that presenting paths of causal factors yields better judgment, and translating these paths into precise probabilistic claims is what we will do here.

In Rescober and Tetlock's analysis [Rescober and Tetlock, 2005], experts receive probability scores for their predictions about events. Using their notation[18], the expert's probability score for an event partitioned into M outcomes is the disparity between the probability assessed to each outcome p_i and an indicator variable assigned to the outcome's occurrence, x_i, normalized by the number of outcomes, or $PS = (\sum_{i=1}^{M}(p_i - x_i)^2)/M$, subject to the constraint that $\sum_{i=1}^{M} p_i = 1$, with the low score of zero being the best, and the high score of one the worst. Almost all experts polled were stymied by a competition consisting of a simple autoregressive model predicting y, where $y[i] = \alpha + \beta_1 y[i-1] + \beta_2 y[i-2] + \gamma_1 x_1[i-1] + \gamma_2 x_2[i-1] + \gamma_3 x_3[i-1]$, where x_1, x_2, and x_3

[18] Or rather, mostly their notation, instead using [] for temporal indexing.

are, among the factors polled, are the three strongest factors by correlation analysis. In other words, what is measured under classical norms are point predictions. In that work, most of these prediction sets refer to a single factor, corresponding well to a structural expression over a single structure in this work. Therefore, we can talk about predictions of s as though they were predictions of Tetlock's x_i without loss of generality. Let us say that Tetlock's probability score, PS, is actually the set of predictions made some time prior using the information available at that time about the time of the event, or $PS(t-n, t-n, t)$. For point predictions, this distinction makes little sense, but observe that if one can make predictions contingent on events yet to occur, then it would be possible to render a different prediction after the time of the prediction, but before the event occurred.

What we would like to do is to develop a causal probability score that reduces to point predictions, but includes causal structures. In other words, specifying a causal model will result in a set of point predictions that can be scored identically to those of Tetlock's analysis, but can be elicited causally. In order to have a causal, instead of a classical model, we need to allow the expert to provide two additional elements:

1. **Dependencies** We would need not only to score point predictions, but would also need to score paths of causal dependencies predicting those factors.
2. **Interventions** We would need to elicit the conditions under which those paths are intervened upon, or severed.

Predictive elicitation methods that allow for conditional dependencies are far from unprecedented. One of the best known is combinatorial prediction markets, which allows for conditional bets to be assembled by trading boolean combinations of combinatorial bets [Hanson, 2007]. It is a matter for future work to determine if the measures suggested here are equivalent.

Before proceeding with an algorithm to generate point predictions from dependencies and interventions, let us describe what we are going to do. For every predicted point, the expert can describe any number of events that could lead to that point, and their likelihood of the event occurring and bringing it about, as well as the likelihood of the event being the case. They can also provide dependencies between events, that sever the effects of another event. If two events together make an outcome more likely, or if two events occurring together might lead to a mixed outcome, it is better to specify a third event based upon this joint precondition. Each point prediction is the average of the independent paths of potential events, with their weights appropriately reduced by severing effects.

By default, the expert should be held accountable for all predictions leading to the predicted point, which is to say all paths of causal dependency. Similarly, we should not hold experts accountable to predictions dependent on events with antecedents that did not occur, except to the degree that the expert specifically predicted those antecedents.

Given this, an algorithm for calculating the point prediction is merely reduce the weight of each path by the maximum degree that is severed, but doing so iteratively to avoid cycles in the severing, and then taking the average of the resulting independent weighted paths (see Appendix G for formalisms related to this section).

Algorithm 2 Building Point Predictions from Event Paths and Interventions

```
Mark all events causing interventions as changed
while At least one dependency is still marked as changing do
  for all Paths where path has changed weight do
    Apply dampened change to weight of path
    if The change of the weight of the path is small then
      Remove change mark on path
    end if
    Calculate the weight of paths the path intervenes on
    if The change of the weight of the intervened path is large then
      Add change mark to intervened path
    end if
  end for
end while
Average the predictions of the resulting disjoint weighted paths
return The resulting point-wise predictions
```

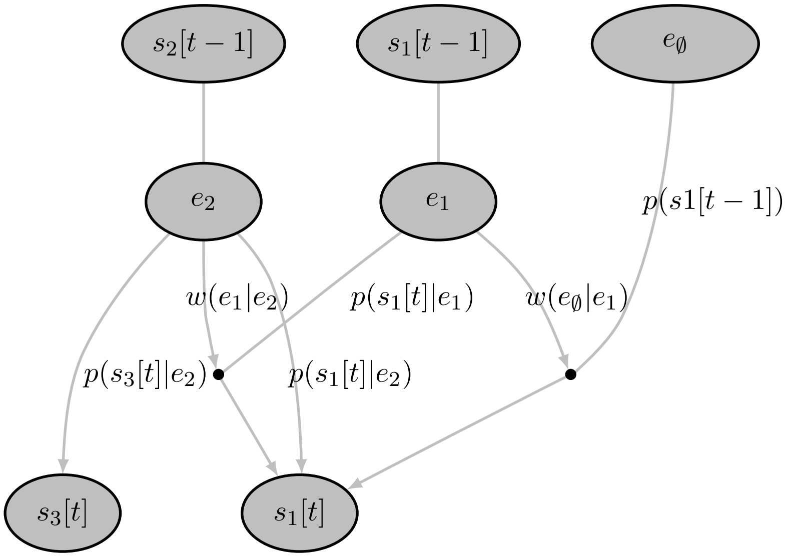


Figure 6: An Example of Event Predictions and Severability

In Tetlock's analysis, he required that a prediction be rendered for every point. A similar requirement is that every point considered to be predicted must either a point prediction or have a path of events and dependencies to predicates that either can be resolved currently or have point predictions of their own. Further, every path of events that cannot currently be resolved would be counted neither for nor against the experts present assessment as far as the $PS(t-n,t-n,t)$ was concerned. However, it would also be possible to compare this probability score against a prediction using all the information known up to the event itself $(PS(t-n,t-1,t))$ or any time in between $(PS(t-n,t-k,t))$. The disparity between the score at the time of the assessment and just before the event could represent how contingent the expert believes the event is, and thus we call it the predicted contingency[19].

In order to see how this works, let us take a look at an example (see Figure 6). Let us consider the likelihood of some structure at time $(p(s_1[t]))$, say that the majority party has had its leader reelected. This probability may change drastically if majority party

[19] The predicted contingency has interesting variants. For example, we could describe the maximum possible disparity between these two scores as the maximum contingency, and the difference between them given what actually occurred as the actual contingency.

was in leadership previously ($s_1[t-1]$) and there was economic prosperity during that time (e_1). If we know about this event, then we would expect that the election result would be contingent upon the economic prosperity ($p(s_1[t]|e_1, s_1[t-1])$), and would say that our prediction independent of this knowledge is now irrelevant ($w(e_\emptyset, e_1)$), which we take as the inverse likelihood). However, at the same time as prosperity, the majority leader may have passed legislation on a controversial moral issue (e_2). Then, we expect that this will cause another candidate taking a leadership position in opposing the legislation to be more likely to be elected ($s_3[t]$ becomes more likely via $p(s_3[t]|e_2)$), make it less likely that that the current majority leader will be elected ($p(s_1[t]|e_2)$), and reduce the effect of the economic prosperity contributing to support ($w(e_1, e_2)$).

Overall, there are any number of modifications to this scheme that might yet be appropriate. If we credit experts for their contingent predictions, then in order to give the expert the best circumstances, we would also need to elicit dependencies that would 'explain away' consequences due to hidden antecedents. If a consequence does not occur despite its antecedent being satisfied, this could offer evidence that other consequences will or will not occur, despite neither being caused or prevented by the non-occurring consequence. This is due to an unknown hidden intervention that has severed known paths of events.

As another example, we might want to allow experts to create predictions with more sophisticated structures. Consider the phrase "I do not know why it's happening, but there seem to be more and more incidents of bankruptcy in Atlanta, and we should take that into account in with all the other factors I've described." In that case, we would want to have different point-wise predictions determined by temporal index instead of any particular event.

However, even through further improvements are possible, we have accomplished what we intended to do, which was to provide a way to elicit causal predictions that accommodate both paths of predictions and severings in those prediction paths while still resolving to point predictions accountable to a Tetlock-style probability score.

Review of Objectives

This section looks at the project as a whole and demonstrates that it accomplishes what it set out to do. First of all, in order to make such an assessment, let us review the objectives.

- Engineer scenario representation methods that allow for the capture, analysis, storage, and reuse of causal and impact information.
- Develop elicitation methods that progressively delimit and arbitrate governance deficits.
- Implement simulation methods capable of demonstrating plausible scenarios from elicited causal structures.
- Position uncertainty discovery as a valid governance need.

Now, let us take a look at each of them individually, to see where in the text that these were met.

- Engineer scenario representation methods that allow for the capture, analysis, storage, and reuse of causal and impact information.

This has been done. A scenario description language given in the section entitled *Structural Elements*, when supplemented by the coding grammar in Appendix A, allows for scenarios to be represented with an open ontology, allowing for new elements to be freely added, while capturing the core notions of how events unfold leading to stakeholder impacts. The results of this representation can be analyzed both through simulation methods and through an equivalent to Tetlock's probability score (as described in the section *The Relation between Classical and Causal Bayesian Norms*).

- Develop elicitation methods that progressively delimit and arbitrate governance deficits.

This has been done. In the section entitled *Modeling Risk Governance Deficits*, we showed that the scenario representation we engineered, when paired with certain methodological safeguards avoiding support biases and attempting comprehensiveness, could help arbitrate risk governance deficits by eliciting paths of risk. The section *Interview-based Discovery Process*, as described in detail in the telephone script in Appendix C, is an elicitation process which gathers the model structures while honoring those methodological safeguards.

- Implement simulation methods capable of demonstrating plausible scenarios from elicited causal structures.

This has been done. Given care, this method can simulate how scenarios will play out from particular points of view. The section *Assembling Structural Elements into Processes* explains how such simulations are undertaken. Having said that, there is certainly no end to the kinds of simulation analysis the could be yet undertaken on these models, and an integration with other systems and methods is an exciting direction for future work.

- Position uncertainty discovery as a valid governance need.

Depending on one's interpretation, this objective has either been met or has just begun. This paper demonstrates that recognized governance deficits may be addressed through discovery processes that find the incompletenesses and implicit uncertainties among the relevant stakeholders. However, positioning not only means to place it in its proper context, but to place it in the view of those who can use it. However, this is not the work of a single project, and I am confident that this project offers a solid foundation for this experimental work, which will demonstrate effectiveness of these methods or their successors to those who can use them appropriately.

Carbon Mitigation Technologies: An Example Problem

Let's now look at carbon mitigation technologies as an example problem. It has all the hallmarks of a distributed risk problem: a very widely distributed common resource with contention about the underlying dynamics (in this case, the atmospheric concentration of carbon dioxide and other greenhouse gasses that can be supported with causing certain levels of harm) and widely varying distribution to whom the harms impact, and also local mitigation measures that may be ineffective and, when taken individually, are likely insignificant.

Given this, what we would like to do is connect the risks of climate change with the costs of their mitigation. We can use the infrastructure development pipeline to discover some of the perspectives of those it would be essential to involve in the initial interview process for the risk governance of a new sequestration-based mitigation technology[20]. These perspectives include those of businesses with carbon liabilities, climate scientists and technologists, technology investors, business investment programs, environmental regulators, insurance companies and insurance regulators, global environmental activists, and individuals in proximity to facilities, as well as carbon market makers and their critics.

Of those, let's look in detail at some of the potential perspectives of three of these stakeholder groups: businesses with carbon liabilities, environmental regulation, and insurance regulation.

[20] Carbon air capture would be a good example, although it is currently facing strong feasibility challenges Ranjana and Herzog [2010].

Potential Perspectives for Businesses with Carbon Liabilities

Not all industries are capable of reducing their emissions. The cement industry, for instance, currently has no proven economical alternative to the highly-emitting process currently used, which is one-for-one in CO_2 emitted per ton of cement produced, with one half of that fundamental to the calcination of $CaCO_3$ [Greer et al., 2000]. Airlines might also be hard-pressed to find substitutes. Despite these challenges, any number of potential viewpoints prevent their willingness to consider sequestration as an option.

Basic viewpoints:

- Sequestration, or perhaps emissions generally, is entirely off of their radar (it's not an airline problem, it's an aerospace engineering problem)
- Thinks of it carbon sequestration as only an alternative for direct emitters (it's for companies that have coal plants) and doesn't think about sponsoring sequestration activities as an alternative.

Technical viewpoints:

- Believe that it is technically infeasible, and that current scientific results are questionable
- Believe that it is technically feasible, but cannot scale

Core business viewpoints:

- Think that it will be more expensive at its final stages as compared to the alternatives that they are invested in
- Think that it will be competitive, but don't want to pay for it now (free-riders)
- Think that it has too risky a business pipeline and that investment won't reward them

- Think that the deployed business will have too many, too severe, or too long-term risks (pollution, leakage, etc)

Competitive viewpoints:

- Have already partnered with companies working on alternative mitigation strategies

Governance viewpoints:

- Thinks that investing in this field won't save them from regulatory penalties for carbon
- Thinks that showing interest may invite early regulatory pressure to adopt

Public relations issues (interested as a business, but not part of public strategy):

- Thinks public will find it to be a cop-out for not working on more front-end reductions
- Thinks public will find political issues in being affiliated with strategies that are not output-reduction or directly biologically/ecologically-driven

Potential Perspectives for Environmental Regulation

One early observation is that current recommendations are being made with respect to a ppm ceiling [Stern, 2007]. Approaches based solely on this number are likely to be problematic, as there may be different risks for concentrations over time, as well as risks due to sudden changes in concentration. Rapidly approaching the ceiling and then leveling off is particularly risky, as other ecological balances may shut down due to steep change or unexpected saturations in mitigating processes [Lenton et al., 2009].

One possible policy approach is liability above a certain concentration (you pay the cost of remediation for what you output above that point), but retroactive liability above a higher concentration (the polluter pays the cost of remediation for what the polluter has output in some substantial percentage of history). There also needs to be a way to make incentives for reducing the overall volume of the carbon overage over time.

One question this brings up is how to make a transition to a correctly priced carbon market when most firms may not be able to pay the current cost of capture and sequestration. A similar problem can be found in the analysis of leaking underground storage tanks (USTs) at gas stations [Boyd and Kunreuther, 1995]. Many gas stations don't have a high profit margin, so the cost of cleaning up after a leak may mean that they have to skimp on upgrades, leading to a greater likelihood of leak in the future, or exit the market entirely, leaving the mess for public cleanup in any case. So, if due care is taken in reporting leaks, then the public might appropriately bear the cost of cleanup, as not to discourage reporting, and to allow for upgrades. Carbon sequestration is also a leaky technology, where its effectiveness needs to be measured in terms of the time-to-leak [Herzog et al., 2003]. Given this, let us see if the analysis of leaky tanks applies to carbon sequestration.

Let us temporarily make the assumption that we can appropriately price costs, previous arguments notwithstanding. There is a cost of sequestering a given volume of carbon to a given depth at a particular time. Importantly, this cost is composed of both a fixed cost related to having the technology, infrastructure (piping, equipment, operational personnel, etc.), and superstructure (monitoring and governance) in place (decreasing over time to a fixed, and still not insubstantial, point), as well as a marginal cost per volume. There is then also percentage of this sequestered carbon leaked at a given time, given when it was sequestered and to what depth. Let us further say that the technology of mass sequestration is largely dependent on geology and can't be engineered to any great extent, such that any carbon sequestered to a particular depth at a particular time will leak at the same rate. We can also talk to about losses due to particular levels of greenhouse carbon, and also say that above some threshold this loss is accelerating due to feedback effects (or, in other words, that its derivative strictly increasing), and that also there are some points where the acceleration accelerates (due to new feedback effects). Let us also take into consideration the total assets of a particular firm with potential carbon liabilities, where we designate the break-even point as zero, below which the firm exits the market. Such a firm would have a particular rate of emissions, and that the damage done by those emissions corresponds to We say that the damage done by a particular firm is the sum of the losses caused by this rate of emissions across time given the greenhouse gas levels at that time, and that therefore the value of sequestering to a given depth is the total loss deferred until leakage.

It would be tempting for a firm to delay cleanup, saying that if they sequester later, the marginal value of the amount sequestered is much higher, as it likely occurs against a larger carbon background. This is deceptively attractive, as that would imply that later efforts are marginally more effective in decreasing risk. However, this discounts any lasting damages from earlier periods, assumes a strictly increasing greenhouse gas background, and ignores any discount factor that might have made it more economical to defer the losses. This also assumes that more stringent penalties aren't assessed the closer one gets to a threshold, as the losses from crossing may be appropriate to merit a retroactive regime even if it causes the market exit of the firm. Nonetheless, delaying may be merited in the case where the leakage from sequestering earlier would have caused a threshold crossing.

Trickier still are capital improvements that reduce the amount of emissions (or, worse yet in terms of analysis, research into capital improvements which may or may not reduce the amount of emissions). However, if we permit ourselves to talk about an expected reduction in emissions given this research, despite how poorly this expectation is formed, we can make the following observations:

- If participants perceive a transition to a retroactive regime as a valid possibility, it may have similar effects to retroactive regulation itself. Unfortunately, this is not all good, as it may withhold credit from firms at critical times, firms who later would have been able to pay for damages.

- It may be that, after a threshold concentration is passed, there is nonetheless a window for the positive feedback loops that cause the threshold to truly take hold. In this case, it would make sense to undertake immediate remediation, by switching to a retroactive regime.

- The expected, temporally discounted improvement caused by an emissions reduction must be greater than the leaky, discounted reduction caused by using some amount of assets to sequester, and applying later profit to the improvement

- A firm that cannot sequester its output will only be viable if an expected improvement will create reduction that can be paid for with later sequestration, even with discounting.

- Policy could do worse than to charge fixed costs to those that can pay for it, and marginal costs to those responsible for the damages.

- One particular problem is that firms will not mitigate if the probability of discovery is so low as to mitigate the expected cost of penalties [Boyd and Kunreuther, 1995]. Therefore, penalties for misreporting should be severe, including jail time for decision makers and executives.

Potential Perspectives for Insurance Regulation

Climate change presents the insurance industry the potential for losses nearly across. This cornucopia of risks include product liability for carbon-intensive products; geological damage to buildings and infrastructure; impacts to public lands; increased risk of respiratory illness and heat mortality; loss of private lands that impact commercial use; decreasing agricultural yields and increased food prices; reduction in fishery populations; mobilization of wastes and post-weather mold; poor financial performance and business failure; interruptions to supply chains, telecommunications, transportation, and operations; disruptions of energy, water, and other public utilities; weather-related increases in commercial and personal vehicle damages; claims of disinformation and professional malfeasance in climate-related decisions; increased need for disaster preparedness and adaptation by third parties; cross-border economic damages from new regulatory and tax exposure; cross-border carbon market policy asymmetry risk; and the risk associated with supply-side technology migration and mitigation measures [Ross et al., 2007].

Despite this wide array of legal liabilities for climate change, the liabilities for carbon sequestration itself are much better understood [de Figueiredo et al., 2005] [de Figueiredo, 2007]. Pumping pressurized carbon-dioxide is already a technique for drawing more oil from wells, legally operational today [Sharp et al., 2009]. Further, the liability of carbon sequestration itself is much better understood today. These risks are summarized in [Leiss, 2009], and very briefly include suffocation due to pipeline or storage breach, induced seismicity, and groundwater contamination. Given that these risks are understood and mirror those in practice today, it seems likely that sequestration research will be able to proceed.

Demonstrating Model Making

Despite having met the objectives we set out to accomplish, the reader may be justified in feeling that they do not have an intuitive understanding of how this approach works. How is it that this technical modeling practice described here actually functions in finding risk governance deficits? Therefore, let me present an example of modeling a few phrases of a document in detail. As we go along, I'll analyze what we learn about the domain, the document, and the modeling formalism. Let me say ahead of time, however, what is reasonable to expect. Documents are in no way subject to the same kind of methodological control as an interview or a computer program designed to ask questions in particular ways. There is no promise that documents will not take any number of assumptions about shared knowledge or worldview for granted, or will actually describe any particular set of situations cohesively. Although writing with clarity and specificity makes it more likely that a reasonably structured worldview is articulated, it is certainly not guaranteed. Even documents that aim for exemplary specificity fall short.

One such exemplary document is "Major Tipping Points in the Earth's Climate System and Consequences for the Insurance Sector", which was mutually produced by Allianz SE and the World Wide Fund for Nature [Lenton et al., 2009]. This document had the advantage of being told from the perspective of both ecological and economic concerns, and also described with specificity the specific tipping points and their expected consequences.

Suppose one tries to represent this phrase: *"A global sea level rise of 0.5 m by 2050 is estimated to increase the value of assets exposed in all 136 port megacities worldwide by a total of $US 25,158 billion to $US 28,213 billion in 2050"*. I ran into two representational problems in the model, both of which resulted in straightforward fixes. First of all, impacts were often described in terms of absolute time. In order to represent this, it was important to make sure that impacts could be parameterized by temporal expressions. It was also necessary to give a model a current date that would serve as a reference point of what defined the present.

The second issue was also straightforward. Although events can change the topology of structures, it is not necessary that they do so. Many events are better represented as change in the values of existing structures. Therefore, I added 'shifts' to the grammar for consequences and implemented the underlying machinery for those transitions. As a result of these modifications, I could write the following model:

```
event globalSeaLevelRise
```

```
      if $time after [time, 2050, 1, 1]
         and $climateChange
              contains [climateChange, continuesUnabated]
         and $seaLevel contains [global, seaLevel]
      then $seaLevel shifts meters=meters+0.500001
      according to AllianzSE, WorldWideFundForNature.

impact portAssetExposureIncreasesInGlobalSeaLevelRise
       3055000000000 usDollarsExposure to insuranceSector
       if (global seaLevel meters:>0.5)
       according to AllianzSE, WorldWideFundForNature.
```

However, the phrase *"A global sea level rise of 0.5 m by 2050 is estimated to increase the value of assets exposed in all 136 port megacities worldwide by a total of $US 25,158 billion to $US 28,213 billion in 2050"* also presented unresolvable interpretive challenges. For example, what happens if the global sea level rises to this 0.5 before or after 2050? This increased asset exposure is caused not only by the rise, but by the change in the assets themselves due to projections of urbanization, and therefore have an unspecified impact. Even more precariously, we do not suppose that a sea-level rise of 0.45 m is going to be without increased asset exposure. A reasonable naive assumption would be to assume a linear increase in exposure, but it's hard to expect that to be true. Instead, we should expect any number of bursts in possible exposure corresponding to the assets of the ports submerged or subjected to maritime weather.

Admittedly, it is unfair to analyze a single sentence closely in isolation. To build an accurate progression of events takes paragraphs, and this sentence illustrates what a good executive summary should do: take you immediately to the conclusion so that you can evaluate for yourself if the chain of events is worth understanding. Nonetheless, this sentence demonstrates exactly what is deficient about point predictions, no matter how accurate: they leave us without the content of the processes that produced them.

Future Work

In many ways, this project raises more questions than it answers. This is both a success and a failing for a project developing analytical methods. One purpose for pursuing formal methods in the first place is so that we may find issues in our previous conceptions. In this regard, this project is a solid success. To review the opportunities resulting from this paper, we first look at follow-on technical developments suggested by the work within the paper. Then, we will look at possible future directions for expanding this work into other domains.

Follow-on Technical Developments

First of all, I look forward to implementing, evaluating, and using the model developed in the section entitled *Assembling Pragmatic Causal Categories*. A suitable language for such an implementation is MIT's Church programming language [Goodman et al., 2008a], which supports the probabilistic programming of non-parametric Bayesian models in a straightforward way. Although Church is currently a self-contained environment, I have every confidence that it can be extended to be an effective utility program.

One important change is to add contingent presents and histories. Right now, we take the current state of affairs to be a simple assertion by the participant, but this is not true. They may not know what happened, but have suspicions. Alternatively, they may be prepared to admit that the histories they've learned may yet be wrong. Therefore, the relationships between structures and their temporal indexes are also subject to likelihoods and dependencies.

Another implementation I will be pursuing immediately is to write a causal, instead of time-series, baseline for Tetlock's sophisticated competition. In this approach, instead of looking to the most immediate preceding factors that could have made an inference, we will look to discover latent common causes between the factors. In short, for each factor identified as a term in Tetlock's model, we will look to see if that in fact was the cause, or if another term in the model, perhaps further back in the series could act as a common cause. This project will also present an opportunity to use the toolkit of known causal discovery algorithms.

There are also any number of possible improvements to the simulator itself. One such improvement includes making the simulation more realistic (such as supporting events that take different durations using an event calendar). Another application includes implementing more algorithms to improve agent learning. Of course, usability and visualization also present key opportunities.

In order for it to come to widespread use, we will have to provide interfaces that practitioners can use easily. These include tools to support running simulations and analyzing the results, as well calculating and reporting on the likelihood of different model categories. Right now, the interviewing application provides one easy way to gather model data, but as for coding it and running analysis methods, one right now needs to work at the level of the code itself. Just as the code requires interfaces, so does the documentation on how to use it. Although this document serves to present the academic merit of the model and as a fine reference for technical purposes, it is not an easy-to-read guide for the practitioner.

Potential Future Directions

This paper proposed an interviewing technique suitable for engaging a distributed set of stakeholders with contradictory worldviews. This method should be undertaken and refined based upon its findings. However, I am not yet sure in what context these interviews will be undertaken.

There are many circumstances where interviewing is not a viable alternative. There are a variety of reasons why one may not be able to conduct an interview in good faith, including stakeholder distrust of the participant's institution, organizational restrictions against interviewing practices, or just a failure to be able to coordinate on a reasonable timescale. In the case where direct elicitation is not possible, it still may be possible to learn a substantial degree about the perspective of a stakeholder from the analysis of open data sources, such as documents posted to the public on the internet. There has already been some very promising work on discovering causal assertions in text (for example, see [Riaz and Girju, 2010]).

However, text analysis is an exceptionally challenging domain. It is itself a very broad and well-studied area, enmeshed in brutally hard research challenges. It is daunting to include a field that includes intelligence analysis, linguistics, library science, computer science, the methodologies for interpreting historical documents, and so forth. Even more challenging than this, however, is the way the documents are provided: they are provided indirectly, with no promises of answering the kinds of questions one wishes to pose, and with no guarantees of having been produced in the kinds of conditions which guard against biases. In this paper, we have already demonstrated that even the most concrete of sentences open up semantic ambiguities that impede a cohesive understanding. For this reason, I am extremely pessimistic about textual analysis methods reaching any kind of resolution of problems such as those posed in risk-governance deficits. However, I do think that document analysis can contribute to an exploratory workflow and could supplement a multi-modal means of inquiry. In particular, I think that questions, similar to those of the interview protocol, can be used to annotate documents, revealing questions that the document does not answer.

Implications

Now that the project has been described completely, let us review the implications of this work. We will see that this work has much to say about using design methods to build regulatory connectivity, how to distribute design methods, the application of new quantitative approaches to design and the implication of these approaches to design thinking, the role of quantitative methods in foresight, and the impact of design methods on the quality of foresight.

This project operated on the assumption that an appropriate design strategy for the mitigation of distributed risk problems is the development of regulatory connectivity. Activities that involve increasing risk, such as living near tropical beaches, need to have the cost of this increasing risk captured and spent to mitigate this growing risk. In particular, insurance that is subsidized both incentivises risky activity and fails to capture the resources that could mitigate those risks, assuring that when catastrophe strikes more will be in the position to be harmed by it, and will have few resources to cope. This project has offered at tool to assist building regulatory connectivity by attempting to infer shared models among disparate stakeholders and any chance that this tool has in doing so is a sufficiently powerful motivating implication for me.

I am very pessimistic at this point that the kind of analysis necessary to fully assess the appropriate regulatory connectivity can be built. Models of this kind will inevitably discover any number of incompatibilities and incompletenesses in the available knowledge, as well as any number of flaws in their own formulations. Yet as pessimistic as I am about any ultimate findings, I am equally as optimistic that improvements in discovery and analysis will make marginal contributions. That is, even scattershot and coarse connections between the costs of harm and the funding of mitigation will have positive impacts. We do not know where this kind of technological forcing may take us, but in the face of uncertainty it is appropriate to hedge. Hedging, and learning from our hedges, is surely as much a part of our exploratory portfolio as design and modeling activities are.

This project demonstrated that design methods can be applied on a continuous and adaptive basis, instead of on the individual meeting, project, or milestone. Furthermore, these methods can be applied with an unknown and changing pool of participants. Such methods, like the problems that they attempt to tackle, are distributed, and are not necessarily linked to a particular place. The method explored here is limited in this capability, as it is strongly constrained to a particular "logical time" due to the constraint of exchangability. Having said that, this approach lights a path for methods more suitable to non-exchangeable processes, while using the same causal abstractions.

Practitioners in design methods should take this project as an early warning sign that a flood of structural and semi-quantitative methods and tools will be developed in the coming years. In particular, a whole new class of statistical methods may be opening up for designers. By being able to use non-parametric Bayesian methods to reason about what we do not know, the designer can use what they have discovered so far to make compelling arguments for further exploratory activity. Additionally, design practitioners who regularly use particular design processes acquire hard-won deep tacit knowledge about the character of the discovery process. It is my suspicion that these designers are implicitly learning inductive constraints appropriate to those methods of inquiry, and it is this learning that explains the discovery power of design, instead of the more generic "generate and evaluate" explanation of design thinking, which explains the capacity of practitioners to learn in a much more limited way. If the generic explanation were true, then uniquely powerful mediums that draw on human capabilities, such as visual thinking and storytelling, should have no special force. Methods like these hope to capture the inductive power of human capabilities.

I was only able to scratch the surface in demonstrating the kinds of tools that will capture capabilities, giving a hypothesis that a particular kind of inductive constraint (causal constraints) might be appropriate for certain kinds of inquiry (open-ended inquiry aimed at sensemaking) in a given stage of risk governance (early stakeholder discovery) for a specific class of problems (those with cross-culturally recognized harms which are likely irreversible in nature).

Over time, foresight and design practitioners alike may find these statistical tools not only aide design processes, but assess the degree to which they managed to discover the relevant criteria effectively. However, as things stand now, I think designers should argue that it is to preliminary to judge the success of design approaches, but instead should be used to demonstrate the success of deciding to design. The question "How is it that the information informing designs is discovered and accumulated by design processes?" should yet challenge statisticians, computer scientists, and psychologists. I personally suspect that the advances in causality, far from being limited to the catastrophic risk domain, will come to serve as conceptual underpinnings for many new design and analysis methods. Introducing causal concepts will help to escape early-stage clustering and categorization, which are blunt instruments for bringing ideas together (a view shared by [Christensen and Raynor, 2003]).

This project also sought to apply the qualitative insights of technical methods to foresight. This is comparatively rare, as the uncertain relationship between forecasting and foresight recapitulate other academic divides between quantitative and qualitative approaches. In the course of this project, I came to a position on the subject. While foresight is distinct from forecasting, the relationship between forecasting and foresight is common-sense: forecasts are source materials for foresight. Any forecast could serve as the source material for a sufficiently discerning and critical foresight scholar. Foresight can hold forecasting to a critical standard by tracking when, and more importantly why, forecasts succeed and fail. Forecasting methods also contain, and fail to contain, concepts about how the future should be represented, and these concepts should be available to foresight scholars.

This project also has implications for the interaction between design and foresight. Design already has a role in prototyping the human changes, allowing us to select more desirable futures. However, design has a further role, in that the design of how we ask changes the content of people's understanding about the future. How we ask people to predict helps determine what they predict, so it is better to ask in ways that help people be nuanced and wise.

Conclusion

This project is a foresight project, and as such examined concepts from which to build possible futures. This project found a rich conceptual material, namely the risks of infrastructure transition, which affect and are affected by almost all aspects of our ordinary lives. We found that these transitions would lead to certain regret, and that any appropriate risk governance policy should strive not only to mitigate that regret, but to survive past failures to continue to offer reasonable guidance. However, what reasonable consists of depends on one's worldview, and so risk governance must concern itself with taking the worldviews of the governed into full consideration.

Foresight has methods for representing multiple worldviews, but there is some question if they are appropriate foundations given our psychology. Therefore, this project looked to psychology and found new conceptual materials lurking in the study of an ancient philosophical problem: causality. Out of this material, we found ways of representing, simulating, modeling, and eliciting the future that could both withstand the most common problems within risk governance and offer better design interfaces for eliciting the most basic of predictive measures. These methods speak to new ways to represent possible futures that offer us the consolation of competence despite inevitable regrets.

References

Abelson, H., Sussman, G. J., and Sussman, J. (1996). *Structure and Interpretation of Computer Programs*. MIT Press, Cambridge, MA, 2nd edition.

Aldous, D. (1985). Exchangeability and related topics. *Ecole dete de probabilites de Saint-Flour*, XIII-1983:1–198.

Allen, J. F. and Ferguson, G. (1994). Actions and events in interval temporal logic. *Journal of Logic and Computation*, 4:531–579.

Bertsekas, D. P. and Tsitsiklis, J. N. (1996). *Neuro-Dynamic Programming*. Athena Scientific, Belmont, Ma.

Bovet, J. and Parr, T. (2008). Antlrworks: an antlr grammar development environment. *Software: Practice and Experience*, 38:1305–1332.

Boyd, J. and Kunreuther, H. (1995). *Retroactive Liability and Future Risk: The Optimal Regulation of Underground Storage Tanks*. Resources for the Future, http://ageconsearch.umn.edu/bitstream/10768/1/ dp960002.pdf.

Braa, K. and Vidgen, R. (1999). Interpretation, intervention, and reduction in the organizational laboratory: a framework for in-context information system research. *Accounting, Management and Information Technology*, 9:25–47.

Brand, S. (2009). *Whole Earth Discipline: An Ecopragmatist Manifesto*. Penguin Books, New York.

Cascio, J. (2010). A survival guide to geoengineering. *Momentum*, pages 23–24. Institute on the Environment, University of Minnisota.

Christensen, C. M. and Raynor, M. E. (2003). *The Innovator's Solution: Creating and Sustaining Successful Growth Businesses*. Harvard Business School Press, Boston, Ma.

Cook, P. (2003). *The City, Seen as a Garden of Ideas*. The Monacelli Press, New York.

Cravens, G. (2007). *Power to Save the World*. Random House, New York.

de Figueiredo, M. A. (2007). *The Liability of Carbon Dioxide Storage*. PhD thesis, Massachusetts Institute of Technology.

de Figueiredo, M. A., Herzog, H. J., and Reiner, D. M. (2005). Framing the long-term liability issue for geologic carbon storage in the united states. *Mitigation and Adaptation Strategies for Global Change*, 10:647–657.

de Mesquita, B. B. (2004). The methodical study of politics. In Shapiro, I., Smith, R. M., and Masoud, T. E., editors, *Problems and Methods in the Study of Politics*, pages 227–47. Cambridge University Press, Cambridge, UK.

Dervin, B. (2001). *What we know about information seeking and use and how research discourse community makes a difference in our knowing*. Health Information Programs Development, National Library of Medicine, Bethesda, MD.

Fontanella, B. J., Campos, C. J., and R., T. E. (2006). Data collection in clinical-qualitative research: use of non-directed interviews with open-ended questions by health professionals. *Rev Lat Am Enfermagem*, 14(5):812–820.

Gill, J. and Walker, L. D. (2005). Elicited priors for bayesian model specifications in political science research. *Journal of Politics*, 67:3:841–872.

Glaser, B. and Strauss, A. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Aldine Transaction.

Glymour, C. (2001). *The Mind's Arrows: Bayes Nets and Graphical Causal Models in Psychology*, page 8. MIT Press, Cambridge, MA.

Goodman, N. D., Mansinghka, V. K., Roy, D. M., Bonawitz, K., and Tenenbaum, J. B. (2008a). Church: a language for generative models. *Uncertainty in Artificial Intelligence*, 22.

Goodman, N. D., Ullman, T. D., and Tenenbaum, J. B. (2008b). Learning a theory of causality. In *Proceedings of the 31st Annual Conference of the Cognitive Science Society*, pages 2188–2193.

Gopnik, A. (2009). *The Philosophical Baby: What Children's Minds Tell Us About Truth, Love, and the Meaning of Life*. Picador, New York.

Gopnik, A., Glymour, C., Sobel, D. M., Schulz, L. E., Kushnir, T., and Danks, D. (2004). A theory of causal learning in children: causal maps and Bayes nets. *Psychological review*, 111(1):3–32.

Gopnik, A. and Tenenbaum, J. (2007). Bayesian networks, bayesian learning and cognitive development. *Developmental Science*, 10(3):281–287.

Graham, J. D., Habegger, B., Cleeland, B., and Florin, M. V. (2009). *Risk Governance Deficits: An analysis and illustration of the most common deficits in risk governance*. International Risk Governance Council, http://irgc.org/IMG/pdf/IRGC_rgd_web_final.pdf.

Greer, W. L., Dougherty, A., and Sweeney, D. M. (2000). *Air Pollution Engineering Manual*, chapter Portland Cement. John Wiley and Sons Inc., 2nd edition.

Griffiths, T. and Ghahramani, Z. (2005). Infinite latent feature models and the indian buffet process. *Advances in Neural Information Processing Systems*, 18(17).

Griffiths, T. L. and Tenenbaum, J. B. (2007). Two proposals for causal grammars. In Gopnik, A. and Schulz, L., editors, *Causal learning: Psychology, philosophy, and computation*, pages 323–346. Oxford University Press, Oxford, England.

Hanson, R. (2007). Logarithmic market scoring rules for modular combinatorial information aggregation. *Journal of Prediction Markets*, 1:3–15.

Hedstrom, P. (2005). *Dissecting the Social: On the Principles of Analytical Sociology*. Cambridge University Press, Cambridge.

Herzog, H., Caldeira, K., and Reilly, J. (2003). An issue of permanence: Assessing the effectiveness of temporary carbon storage. *Climatic Change*, 59:3:293–310.

Holladay, J. S. and Schwartz, J. A. (2010). *Flooding the Market: The Distributional Consequences of the NFIP*. Institute for Policy Integrity: New York University School of Law, http://policyintegrity.org/files/publications/FloodingtheMarket.pdf, policy brief no. 7 edition.

Homer-Dixon, T. (1995). The ingenuity gap: Can poor countries adapt to resource scarcity? *Population and Development Review*, 21(3):587–612.

Jarratt, D. (1996). A comparison of two alternative interviewing techniques used within an integrated research design: a case study in outshopping using semistructured and non-directed interviewing techniques. *Marketing Intelligence and Planning*, 14/6:6–15.

Jones, P. H. (2007). Socializing a knowledge strategy. In Abou-Zeid, editor, *Knowledge Management and Business Strategies: Theoretical Frameworks and Empirical Research*, pages 134–164. Idea Group, Hershey, PA.

Jordan, M. I. (2010). Bayesian nonparametric learning: Expressive priors for intelligent systems. In *Heuristics, Probability, and Causality: A Tribute to Judea Pearl*, chapter 10. College Publications, http://www.collegepublications.co.uk.

Karp, R. M. (1972). Reducibility among combinatorial problems. In Miller, R. E. and Thatcher, J. W., editors, *Complexity of Computer Computations*, page 85–103. Plenum, New York.

Kemp, C., Goodman, N. D., and Tenenbaum, J. B. (2010). Learning to learn causal models. *Cognitive Science*, 34 (7):1185–1243.

Kemp, C., Perfors, A., and Tenenbaum, J. B. (2007). Learning overhypotheses with hierarchical bayesian models. *Developmental Psychology*, 10:3:307–321.

Kemp, C. and Tenenbaum, J. B. (2008). The discovery of structural form. *Proceedings of the National Academy of Sciences of the United States of America*, 105(31):10687–10692.

Kemp, C., Tenenbaum, J. B., Griffiths, T. L., Yamada, T., and Ueda, N. (2006). Learning systems of concepts with an infinite relational model. In *Proceedings of the Twenty-First National Conference on Artificial Intelligence*.

Krynski, T. R. and Tenenbaum, J. B. (2007). The role of causality in judgment under uncertainty. *Journal of Experimental Psychology*, 136:430–450.

Kuhnert, P. M., Martin, T. G., and Griffiths, S. P. (2010). A guide to eliciting and using expert knowledge in bayesian ecological models. *Ecology Letters*, 13:900–914.

Kuniavsky, M. (2002). *Nondirected Interviews: How to Get More Out of Your Research Questions*. Adaptive Path, http://www.adaptivepath.com/ideas/essays/archives/000041.php.

LaValle, S. (2006). *Planning Algorithms*. Cambridge University Press, New York.

Law, A. M. and Kelton, W. D. (2000). *Simulation, Modeling, and Analysis*. McGraw Hill, 3rd edition.

Leiss, W. (2009). *Risk Management of Carbon Capture and Storage: Overview and Future Steps*. Institute for Sustainable Energy, Environment, and Economy (ISEEE), http://iseee.ca/files/iseee/LeissCCS-RMfinal.pdf.

Lenton, T., Footitt, A., and Dlugolecki, A. (2009). *Major Tipping Points in the Earth's Climate System and Consequences for the Insurance Sector*. Allianz SE and the World Wide Fund for Nature, http://knowledge.allianz.com/nopi_downloads/ downloads/TP_Final_report.pdf.

Leveson, N. G. (1995). *Safeware: System Safety and Computers*. Addison Wesley, Reading, MA.

Lewis, D. (1973). *Counterfactuals*. Harvard University Press, Cambridge, MA.

Linstone, H. A. and Turoff, M. (1975). *The Delphi Method: Techniques and Application*. Addison-Wesley, Reading, MA.

List, D. (2004). Multiple pasts, converging presents, and alternative futures. *Futures*, 36:23–43.

McDonough, W. and Braungart, M. (2002). *Cradle to Cradle: Remaking the Way We Make Things*. North Point Press, New York.

Miller, J. H. and Page, S. E. (2007a). *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*, chapter 3: Modeling. Princeton University Press, Princeton, New Jersey.

Miller, J. H. and Page, S. E. (2007b). *Complex Adaptive Systems: An Introduction to Computational Models of Social Life*, chapter 5: Computation as Theory. Princeton University Press, Princeton, New Jersey.

Miller, K. T., Griffiths, T. L., and Jordan, M. I. (2008). The phylogenetic indian buffet process: A non-exchangeable nonparametric prior for latent features. *Advances in Neural Information Processing Systems*, 18.

Norman, D. (2010). *Why Design Education Must Change*. http://www.core77.com/ blog/columns/ why_design_education_must_change_17993.asp.

Norris, J. R. (1997). *Markov Chains*. Cambridge University Press, New York, NY.

Onstad, C. (2009). *The Moonwalk Has Come to a Full and Complete Stop*. Achewood, http://achewood.com/index.php?date=06282009.

Orlikowski, W. (1992). The duality of technology: Rethinking the concept of structure in organizations. *Organization Science*, 3(3):398–427.

Parr, T. (2007). *The Definitive ANTLR Reference: Building Domain-Specific Languages (Pragmatic Programmers)*. Pragmatic Bookshelf.

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge University Press, New York.

Pennefather, P. and Jones, P. (2008). Interpretive collaborative review: Enabling multi-perspectival dialogues to generate collaborative assignments of relevance to information resources in a dedicated problem domain. In *Proceedings of ELPUB2008, the 12th International Conference on Electronic Publishing*, Toronto.

Pennefather, P. and Jones, P. (2009). Humbled and interpreting the significance of health science research: A framework for collaborative assessment of sensemaking significance. Draft in Personal Communication.

Pochampally, K. K., Nukala, S., and Gupta, S. M. (2009). *Strategic Planning Models for Reverse and Closed-Loop Supply Chains*. CRC Press, Boca Raton, FL.

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

Ranjana, M. and Herzog, H. J. (2010). Feasibility of air capture. In *10th International Conference on Greenhouse Gas Control Technologies*, Amsterdam, Netherlands.

Renn, O. (2008). *Risk Governance: Coping with Uncertainty in a Complex World*. Earthscan, London.

Rescober, P. and Tetlock, P. (2005). *Expert Political Judgment: How Good Is It? How Can We Know?*, page Technical Appendix. Princeton University Press, Princeton, New Jersey.

Riaz, M. and Girju, R. (2010). Another look at causality: Discovering scenario-specific contingency relationships with no supervision. In *Proceedings of the Fourth IEEE International Conference on Semantic Computing*, pages 361–368, Pittsburgh, PA.

Rincon, P. (2011). *Volcanic ash air shutdown the 'right' decision*. BBC News Website, http://www.bbc.co.uk/news/science-environment-13161056.

Rosenau, J. N. (1992). Governance, order, and change in world politics. In *Governance without Government: Order and Change in World Politics*, pages 1–29. Cambridge University Press, Cambridge, UK.

Ross, C., Mills, E., and Hecht, S. B. (2007). Limiting liability in the greenhouse: Insurance risk-management strategies in the context of global climate change. *Stanford Evironmental Law Journal and the Stanford Journal of International Law*, 26A/43A:251–334.

Schwartz, P. (1991). *The Art of the Long View: Planning for the Future in an Uncertain World*. Currency Doubleday, New York.

Sharp, J. D., Jaccard, M. K., and Keith, D. W. (2009). Anticipating public attitudes toward underground co_2 storage. *International Journal of Greenhouse Gas Control*, 3:641–651.

Snowden, D. J. (2005). Multi-ontology sense making: a new simplicity in decision making. *Management Today Yearbook*.

Spirtes, P., Glymour, C., and Scheines, R. (2000). *Causation, Prediction, and Search*. MIT Press, Cambridge, MA, 2nd edition.

Squires, S. (2002). *American Breakfast & the Mother-in-Law: How an Anthropologist Created Go-Gurt*. National Associations for the Practice of Anthropology, http://practicinganthropology.org/stories/2002/ american-breakfast-the-mother-in-law-how-an-anthropologist-created-go-gurt/.

Stern, N. (2007). Stern review on the economics of climate change. *Journal of Economic Literature*, 7:233–702.

Strauss, A. L. and Corbin, J. (2008). *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. Sage Publications, Thousand Oaks, CA, 3rd edition.

Sunstein, C. (2007). *Worst-case Scenarios*. Harvard University Press.

Sutton, R. S. and Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, Ma.

Taleb, N. N. (2011). *The Future Has Thicker Tails than the Past: Model Error As Branching Counterfactuals*. Social Science Research Network, http://ssrn.com/ abstract=1850428.

Team, S. G. S. (2005). *Shell Global Scenarios to 2025*. Shell Visual Media Services, Shell International Limited, London.

Tenenbaum, J. B. and Griffiths, T. L. (2003). Theory-based causal inference. In Becker, S., Thrun, S., and Obermayer, K., editors, *Advances in Neural Information Processing Systems 15*, pages 35–42. MIT Press, Cambridge, MA.

Tetlock, P. (2005a). *Expert Political Judgment: How Good Is It? How Can We Know?* Princeton University Press, Princeton, New Jersey.

Tetlock, P. (2005b). *Expert Political Judgment: How Good Is It? How Can We Know?*, page 189. Princeton University Press, Princeton, New Jersey.

Tetlock, P. (2007). *Why Foxes Are Better Forecasters Than Hedgehogs*. The Long Now Foundation, http://www.longnow.org/ seminars/02007/jan/26/ why-foxes-are-better-forecasters-than-hedgehogs/.

Thibaux, R. and Jordan, M. I. (2007). Hierarchical beta processes and the indian buffet process. this volume. In Meila, M. and Shen, X., editors, *Proceedings of the 11th International Conference on Artificial Intelligence and Statistics*, volume 11, Madison, WI. Omnipress.

Tilly, C. (2004). Observations of social processes and their formal representations. *Sociological Theory*, 22:595–602.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185:1124–1131.

Verdoux, P. (2010). Risk mysterianism and cognitive boosters. *Journal of Futures Studies*, 15:1–20.

Wolf, K. D. (2002). Contextualizing normative standards for legitimate governance beyond the state. In Grote, J. R. and Gbikpi, B., editors, *Participatory Governance: Political and Societal Implications*, pages 35–50. Leske and Budrich, Opladen, Germany.

Wolf, K. D. (2005). Private actors and the legitimacy of governance beyond the state: Conceptional outlines and empirical explorations. In Benz, A. and Popadopoulos, I., editors, *Governance and Democratic Legitimacy: Transnational, European, and Multi-Level Issues*, pages 200–227. Routledge, London.

Yu, C. (2010). *How to Live Safely in a Science Fictional Universe: A Novel*, page 17. Pantheon Books, New York.

A: Coding Grammar

> *"...a computer language is not just a way of getting a computer to perform operations but rather that it is a novel formal medium for expressing ideas about methodology."*
> -from *"The Structure and Interpretation of Computer Programs"* by Harold Abelson, Gerald Jay Sussman, and Julie Sussman [Abelson et al., 1996]

This grammar uses ANTLR notation [Parr, 2007] and was developed in ANTLRWorks [Bovet and Parr, 2008].

```
model : 'model' ('current' absolutetime)?
      (structure|observation|stakeholder|sense|action
      |event|impact|deference|anticipate|criteria|dependence)+
```

Our model consists of one or more structures, observations, stakeholders, sensings, actions, events, impacts, deferences, anticipations, criteria, and dependences. It may be specified in reference to a particular time, taken as the time when the model was elicited.

```
absolutetime : 'absolutetime' '[' (integerstring)+ ']'
```

An absolute time is a date given by a sequence of one or more decreasingly large calendar units, where the first corresponding absolute time is assumed. For example, `absolutetime [2005 11]` would be November 1, 2005, midnight.

```
criteria : 'criteria' criteriaName=STRING (desc=description)?
                      ('minimize'|'maximize') '.'
```

A criteria is identified by a string, and a description can optionally be provided. Criteria either correspond to rewards and other positive notions, or losses and other negative notions, and thus should be maximized or minimized, respectively.

```
structure : 'structure' stateName=STRING (structuredcloud)?
            (description)? ('current' stakeholderClause )? '.'
```

A structure is identified by a name, and a description can optionally be provided. Structures correspond to various aspects of overall states-of-affairs that are, could be, or could have been. Structures can be as simple as their name, or can be described by more complicated structures called structured clouds. Structures can be designated as currently the case according to one or more stakeholders.

```
observation : 'observation' observationName=STRING
              (desc=description)? ('current' stakeholderClause)? '.'
```

An observation is identified by a name, and a description can optionally be provided. Observations describe aspects of the world which could be directly observable. Observations can be designated as currently the case according to one or more stakeholders.

```
impact : 'impact' impactName=STRING (desc=description)?
         magnitude crit=STRING
         'to' stakeholderName=STRING 'if' ce=cloudexpression
         ('when' te=temporalExpression)?
         stakeholderClause '.'
```

An impact is some magnitude along a particular criteria felt by a stakeholder under particular conditions (where such descriptions would logically match a cloud expression). This may not be their own assessment, but the assessment of another stakeholder.

```
temporalExpression : temporalOp=STRING absolutetime
```

A temporal expression is a temporal operator compared against an absolute time.

```
magnitude : (qualMagnitude|weight)
```

Magnitudes can be expressed qualitatively or quantitatively.

```
stakeholder : 'stakeholder' stakeholderName=STRING (structuredcloud)?
              (description)? ('current' stakeholderClause )? '.'
```

A stakeholder is an individual in the overall state of affairs that may be impacted by events and have assertions and opinions about what is the case.

```
action : 'action' actionName=STRING  (desc=description)?
         ('typically' duration)? '.'
```

An action is something a stakeholder can do, sometimes with a characteristic duration which will carry to events the action is undertaken within, unless otherwise specified.

```
sense : 'observe' name=STRING observationName=STRING 'when'
        stateName=STRING ('with' prob=likelihood)?
        'according' 'to' stakeholderName=STRING '.'
```

A particular observation is sensed when a given state occurs, possibly only with a given likelihood, according to a particular stakeholder. The default likelihood for all units is almost certain, reflecting the convention of speaking deterministically unless otherwise qualified.

```
event : 'event' eventName=STRING  'if' orc=orclause
        'then' var c1=consequence ('and'  var c2=consequence )*
         ('with' prob=likelihood )?
         'according' 'to' stakeholderName=STRING
                    (',' skName2=STRING)*
        '.'
```

If some events happen or stakeholders undertake actions, given the current state of affairs, then some new state of affairs comes into effect, possibly only with a given likelihood, according to a particular stakeholder.

```
dependence : 'depending' name=STRING 'on'
             (compositeVaryingTerm=jointDependenceTerm
                  | singleVaryingTerm=STRING)
     'dependent' (compositeVaryingTerm=jointDependenceTerm
                              | singleVaryingTerm=STRING)
      ('mutually')? ('exclusive'|'independent' | likelihood) '.'
```

For all units for which likelihood may be expressed, it may also be appropriate to express a conditional likelihood between them and units of a similar type. These conditional likelihoods can be mutually exclusive, mutually independent, or dependent to a given degree. The current default is mutual independence.

```
jointDependenceTerm : 'joint' '[' term=STRING
                      (',' term2=STRING)* ']'
```

Joint terms reflect a composite condition over which conditional likelihoods can be expressed.

```
consequence : (('becomes' result=structuredcloud
                ('from' from=structuredcloud )? )
               | 'stops' 'being' result2=structuredcloud
               | 'shifts' cloudMathExpr (',' cloudMathExpr)*)
              ('within' duration)?
```

Consequences cause structures to transition from one state of affairs to another, usually as the result of an event. These consequences can optionally take some duration.

```
cloudMathExpr : tag=STRING '=' argtag=STRING
                ('+'|'-'|'*') value=weight
```

Cloud math expressions take the weight of a argument tag from a structure, change it according to some mathematical expression, and set the value of a tag in the structure with that new value.

```
duration: (weight calendarUnits ('varying' duration)?)
          | calendarUnits
```

A duration is a quantitative calendar unit, possibly with another duration as a usual variation.

```
calendarUnits : ('nanoseconds'|'milliseconds'|'seconds'|'minutes'
                |'hours'|'days'|'years'|'decades'|'centuries'
                |'millenia')
```

Calendar units are conventional measures of time.

```
structuredcloud : cloudchild
```

A structured cloud is a cloud-child with no parents.

```
cloudchild : '(' (wt)* (cloudchild )*  ')'
```

A cloud child is a set of weighted tags and a set of cloud children.

```
wt : STRING (':' weight )?

weight : weightstring

weightstring : ('-')?
       (((INTEGER )+ ('.'  (INTEGER )* )?)
        | ('.'  (INTEGER)+ ))
```

Weighted tags are pairs of names with numerical values.

```
orclause : andclause ('or' andclause)*

andclause : notclause ('and' notclause)*

notclause : 'not' notclause | parenclause

parenclause : '(' orclause ')' | condition
              | 'True' | 'False'
```

Boolean expressions are supported with precedence given to parenthesis, then logical *not*, then logical *and*, and finally logical *or*.

```
condition : conditionalvariable STRING
            '[' (value (',' value)* )? ']'

conditionalvariable  : VARIABLE

value : var  | constant
```

```
constant : STRING

var : VARIABLE
```

Conditions are predicates on variables and constants.

```
cloudexpression : '(' (var
                      | cloudexpression
                      | STRING)* ')'
```

Cloud expressions are logical conditions applicable to structured clouds.

```
deference : 'defer' 'to' knowledgeHolder=STRING
            'on' phenomena=STRING stakeholderClause
            ('with' prob=likelihood)? '.'
```

If a given state, observation, event, or action has taken place, one stakeholder trusts or distrusts the assessment of another stakeholder.

```
anticipate :
    'anticipate' name=STRING anticipatedEvent
    ('and' anticipatedEvent)*
    ('with' prob=likelihood)?
    ('when' consequence)?
    stakeholderClause '.'

anticipatedEvent :
        'event' eventName=STRING   'happens'
        | 'stakeholder' stakeholderName=STRING
          'will' actionName=STRING
```

Some events will happen or some stakeholders undertake actions with a given likelihood if a given condition holds, according to a particular stakeholder or stakeholders.

```
stakeholderClause : 'according' 'to'
                    stakeholderName=STRING (',' STRING)*
```

Many constructs are according to the particular testimony of a stakeholder or stakeholders.

```
likelihood : qualProb|quantProb|'unknown'
```

The assessment of likelihood can either be qualitative or quantitative.

```
qualMagnitude : ('no'|'insignificant'|'low'|'moderate'
                     |'high'|'extreme')?
                 ('declining'|'rising')
```

The magnitude and direction of an impact can be expressed in a qualitative way.

```
qualProb : ('extremely'|'very'|'somewhat'|'moderately')?
           ('impossible'|'low'|'moderate'|'high'|'certain'
                        |'almost' ('certain'|'impossible'))
```

Probabilities and related quantities can be expressed in a qualitative way. Almost certain is reserved for if it is ontologically possible or impossible, but the observer is certain it will not occur.

```
quantProb : 'between' weight 'and' weight
            | ('about'|'near'|'exactly') weight
```

Probabilities can be given by a range or approximated as near a particular value.

```
description :  '"' (STRING|punct)+ '"'
```

Descriptions are double-quoted strings.

```
COMMENT :  '/*' (options {greedy=false;} : . )* '*/'
```

Additional comments, not included with in the model, can be included in the model text if surrounded by C-style comment delimiters.

```
num  : ('-')? INTEGER+ ('.' INTEGER+)?

VARIABLE :'$' ('A'..'Z'|'a'..'z'|'0'..'9')+
INTEGER : ’0’..’9’

STRING : (’A’..’Z’|’a’..’z’)(’A’..’Z’|’a’..’z’|INTEGER|’-’|’_’)*

punct  : (’.’ |’,’ )

WHITESPACE : (’ ’|’\t’|’\r’|’\n’) {$channel=HIDDEN; }
```

Numbers, strings, and punctuation are defined in terms of particular allowed characters. Whitespace characters help delimit tokens but are otherwise insignificant.

B: Notational Shorthand Examples

Structure to possible observation: $structure \xrightarrow{o} observation$

Structure to stakeholders: $structure \xrightarrow{p} stakeholder$

Structure to potential actions: $structure \xrightarrow{a} action$

Stakeholder to potential actions: $participant \xrightarrow{a} action$

Stakeholder to criteria: $participant \xrightarrow{c} criteria$

Actions to potential outcomes: $action \xrightarrow{s} structure$

C: Telephone Script

This section includes a complete example of how to conduct an elicitation session in the domain of climate change and greenhouse gas emissions. The following script uses the convention of *italics* to represent information specific to the interview, such as statements the interviewee has already made in the course of the interview, the interviewee's specific role or profession[21], and question number.

[21] Interviewee role only referenced in pre-interview topics and in grounding questions

Greeting: "Hi. I'm *interviewer name*. I'm *role played within study*. Am I speaking with *participant name*? Good. Previously we had discussed having an interview at this time for my study "Addressing Risk Governance Deficits through Scenario Modeling Practices". Does that still work for you? *wait for response* Excellent."

Disclaimer: "Thank you for agreeing to take part in this study. If you wish, you may decline to answer any questions or participate in any component of the study. Further, you may decide to withdraw from this study at any time."

Instructions: "I'm going to be asking you a number of questions. During this interview, I will frequently ask you follow-on questions about answers you have given previously, sometimes quite a long time ago in the interview. In order to make this easier for both of us, I will give each question, or occasionally a few questions, a number, so that we have a common reference. In our email correspondence, you have been sent a form of numbered blanks. Have you either printed that form, or do you have another way you can take numbered notes for yourself?" (If no, have them get some blank paper or open a word processing program and make numbered entries.)

For each question, first say "question number *number*.

First prompt question:

> Consider climate change broadly, including its ongoing causes, the effects it is causing or is progressing toward causing, and the policies that are being developed as a result. What does climate change look like today?

Second prompt question:

> What is the current involvement of people in *area of expertise* with mitigation technologies that tackle mitigating greenhouse gasses directly, such as carbon sequestration, and how might that involvement change?

Generic prefix for all non-prompt questions: Now, going back to question *number, where*

I have you saying participant answer,

If the questioning takes the full hour: "It's been an hour and I'm mindful of your time, so let us wrap this up."

If questioning is completed before end of hour: "That's all the questions I have."

Revision: "Are there any of your previous answers you'd like to make additional comments about or revisions to?"

Closing: "Thanks for participating in this study. I hope to provide a report of my results by *report deadline*. Have a great day. Bye."

D: Mathematical Notation

$\in$ indicates set membership, or 'is one of these', such that fluffy $\in$ Cats.

$\subset$ indicates set subset, or 'all of these are instances of those', such that Cats $\subset$ Animals.

Curly braces indicate a set that is specifically given, such that fluffy $\in$ {fluffy,garfield,whiskers} $\subset$ Cats

$P(V)$ is a probability distribution over the members of a set, V. For example, if $V = \{0, 1\}$, then $P(V)$ might be the Bernoulli distribution with $p = 0.5$, in which case a draw from $P(V)$ would produce 0 or 1 with equal odds.

$\mathcal{P}(V)$ is the power-set of V, i.e. all sets that can be formed from combinations of the set's members, such that if $V = \{0, 1\}$, then $\mathcal{P}(V) = \{\{\}, \{0\}, \{1\}, \{0, 1\}\}$.

∞ is infinity, used for cases that become arbitrarily large in magnitude.

; separates members of a sequence, while , indicates sequence elements that are proceeding concurrently

x^* indicates an arbitrary sequence of members in a set (in this case X)

$\rightarrow$ is a rewrite relation, or 'changes thing on the left to the thing on the right', except when used in limits.

$\mapsto$ is the same thing as $\rightarrow$, only for specific cases, instead of as a general rule.

lim is for limits, which indicate mathematical terms becoming arbitrarily close; for example $\lim_{i \rightarrow \infty} 1 + \frac{1}{i} = 1$, as the fractional term becomes arbitrarily small, when the denominator becomes arbitrarily large.

$\times$ means binary relation, or 'these two sets taken together as a unit'.

$\forall$ means for all instances of that class, while $\exists$ indicates that there is at least a single instance of that class.

$\wedge$ means 'and'.

$\mathbb{N}$ is the set of natural numbers, i.e. $\{1, 2, 3, \ldots\}$

$\Re$ is the set of real numbers, i.e. the numbers that can be represented with arbitrarily long decimal sequences

$\sum$ is addition over some number of terms, which is usually taken over sequences of integers $\sum_{i=1}^{5} i^2$ is the sum of the first five squares (i.e. $1^2 + 2^2 + 3^2 + 4^2 + 5^2 = 55$), but is sometimes taken over sets, as in $\sum_{i}^{\{1,5,8\}} i^2 = 1^2 + 5^2 + 8^2 = 90$

$\prod$ is the equivalent to $\sum$ for multiplication.

$b|c$ is b conditional on c, or the state of b given that c is the case.

$\perp\!\!\!\perp$ is independence, such that $b \perp\!\!\!\perp c$ means that knowledge of the state of b gives no information about the state of c, and vice-versa

$\sim$ either means "distributed according to" or "similar in some way", depending on context

E: Assembling Structural Elements into Simulable Processes

This appendix provides formalisms for the content discussed in the section entitled "The Challenge of Loss Over Time".

Imagine the set of sequences of complete temporally-confined states-of-affairs, $\ldots; X[-1]$; $X[0]$; $X[1]; \ldots$, where each indexed state is separated by a distinguishable change, where we interpret zero as the current or present state-of-affairs, and where the index denotes some distinguishable change. We can call such a sequence a state-indexed timeline, $\overleftrightarrow{X}$, where a particular timelines over state can be represented as $\ldots; x[-1]; x[0]; x[1]; \ldots$, such that $\overleftrightarrow{x} \in \overleftrightarrow{X}$.

The complete state-of-affairs is a curious structure to build timelines over, as it does not allow any other phenomena to proceed concurrently. However, if we are looking at a more structural level, then we can experience concurrent effects, where multiple changes occurred at effectively the same time. In this case, it makes sense to define the set of possible timelines over the power-set of structures, such that $\overleftrightarrow{s_{\mathcal{P}}} \in \overleftrightarrow{\mathcal{P}(S)} = \ldots; \mathcal{P}(S)[-1]; \mathcal{P}(S)[0]; \mathcal{P}(S)[1]; \ldots$. Given timelines, we can talk about futures $\overrightarrow{\mathcal{P}(s)} \in \overrightarrow{\mathcal{P}(S)} = \mathcal{P}(S)[0]; \mathcal{P}(S)[1]; \mathcal{P}(S)[2]; \ldots$ and histories: $\overleftarrow{\mathcal{P}(S)} = \ldots; \mathcal{P}(S)[-2]; \mathcal{P}(S)[-1]$. We can also talk about relative timelines, as defined over an indexing interval in time, i, such that a relative timeline is constructed as $\overleftrightarrow{s_{\mathcal{P}}[i]} \in \overleftrightarrow{\mathcal{P}(S)} = \ldots; \mathcal{P}(S[i-1]); \mathcal{P}(S)[i]; \mathcal{P}(S)[i+1]; \ldots$.

Further, we can also talk about relative futures $\overrightarrow{\mathcal{P}(S)[i]} = \mathcal{P}(S)[i]; \mathcal{P}(S)[i+1]; \ldots$, as well as relative histories: $\overleftarrow{\mathcal{P}(S)[i]} = \ldots; \mathcal{P}(S)[i-2]; \mathcal{P}(S)[i-1]$. We can also talk about doubly-indexed segments of timelines, or time-slices, such that $s_{\mathcal{P}}[i,j] \in \mathcal{P}(S)[i,j] = \mathcal{P}(S[i]); \mathcal{P}(S)[i+i]; \ldots; \mathcal{P}(S)[j-1]; \mathcal{P}(S)[j]$. Let us designate some arbitrary subsequence or timeslice of structure timelines as $s^*_{\mathcal{P}} \in \mathcal{P}(S)^*$, in loose analogy to the closure operator of automata theory. This notation also includes empty timelines, which will be useful for indicating that nothing has happened in the timeline.

Timelines, futures, histories, and other structures need not only apply to structures. We can also refer to the timelines of events $\overleftrightarrow{\mathcal{P}(E)}$, or of impacts to particular stakeholders $\overleftrightarrow{\Re^{c}_{sk}}$, or of any other aspect of the model which can be said to happen within a particular time.

We can say that the present value to a stakeholder for being in a given state-of-affairs is the sum of the value of all structures that hold in that state, or $r(x[i], sk) = \sum_{s}^{x[i]} r(s, sk)$. However, this vector quantity, although already laden with trade-offs across different criteria, is insufficient for making judgments. Instead, they should then expect, on average, to experience some $E[\overrightarrow{r(x[i], sk)}]$, which is recursively defined as $E[\overrightarrow{r(x[i], sk)}] = r(x[t]) + f_a\Big(\sum_{x[t+1]} \overrightarrow{r(x[t+1])} P(x[t+1]|x[t], a)\Big)$ over future conditions, where f_a is the function by which they actually do end up selecting actions. In the worst case will experience a minimum reward (i.e. a maximum loss) of $min(\overrightarrow{r(x[i], sk)}) = r(x[t]) + f_a\Big(\min_{\overrightarrow{r(x[t+1])}|x[t],a} (\overrightarrow{r(x[t+1])})\Big)$. These quantities may diverge substantially from either what they would expect given their anticipation of how they will act $a = an(x[i+1], sk) \in An_{sk}$, the anticipation of any other stakeholder for how they should act $a = an(x[i+1], sk_1) \in An_{sk_2}$, or the actions that would actually best increase reward and decrease risk, $E[\overrightarrow{r^*(x[t], sk)}] = r(x[t], sk) + max_a\Big(\sum_{x[t+1]} r^*(x[t+1]) P(x[t+1]|x[t], a)\Big)$. Given this formulation, we are now within the realm of planning algorithms [LaValle, 2006], which much of this section references.

Oftentimes, when interested in an arbitrary sequence of events, we will not be interested in the fact that events could be occurring concurrently, but merely that there exists a sequence of events. In that case, we will talk about e^* instead of $e^*_{\mathcal{P}}$. This is not to neglect that concurrency can occur, but merely to simplify the notation. Similarly, we might not talk about a particular stakeholder, but a stakeholder population, which is a sequence of overlapping set of stakeholders holding roughly similar interests (sk^*).

Given this, we can express the average reward for a stakeholder group as the average of the sum of temporal indexes: $E[\overrightarrow{r^*(x[t], sk^*)}]_t = \lim_{T \to \infty} \frac{\sum_{i=t}^{T} E[\overrightarrow{r(x[i], sk^*)}]}{T}$

Alternatively, one can also talk about the discounted reward instead of the overall average using a discount parameter, $\alpha \in (0, 1)$, which geometrically reduces the estimates of rewards and losses as time goes on. Given this, one can describe a discounted expected value of a given state-of-affairs to a stakeholder population:

$$E[\overrightarrow{r^*(x[t], sk^*)}] = r(x[t], sk) + \alpha f_a\Big(\sum_{x[t+1]} R^*(x[t+1]) P(x[t+1]|x[t], a)\Big)$$

Algorithm 3 Scenario Model Simulation Run (Elaborated)

```
(Get the initial conditions)
$S_{sim}[0] = sample(\mathcal{P}(S[0]))$
$t = 0$
(Keep going while the possible absolute discounted risk is greater than some small quantity of indifference for any criteria)
while $\exists C, \alpha_c^t \sum_{r}^{R_c} |value(r)| > \epsilon_c$ do
    (Evaluate the rewards and losses for each stakeholder)
    $R_{sim}[t] = R(S_{sim}[t], Sk_{sim}[t], C)$
    (Sample the actions that stakeholders will take)
    $\{Sk, A, Se\}[t] = sample(An(S_{sim}[t]))$
    (Sample the events that result from the current state and those actions)
    $E_{sim}[t] = sample(E(\{Sk, A, Se\}, S_{sim})[t])$
    (Based on those events, determine the resulting state-of-affairs)
    $S_{sim}[t+1] = E_{sim}(S_{sim})[t]$
    $t = t + 1$
end while
(Give the user everything that happened in the simulation)
return $\overrightarrow{S_{sim}[0]}, \overrightarrow{R_{sim}[0]}, \overrightarrow{\{Sk, A, Se\}[0]}, \overrightarrow{E_{sim}[0]}$
```

F: A Non-Parametric Scenario Discovery Model

This section provides the mathematics for the section described in "Assembling Pragmatic Causal Categories". Let us begin by looking at the complete picture of the model presented here (see Figure 7). For each model element (say the set of elicited stakeholders, Sk), we say that it was generated from some unknown set of true categories for that element that we are sampling, which we will indicate with overline notation ($\overline{Sk}$).

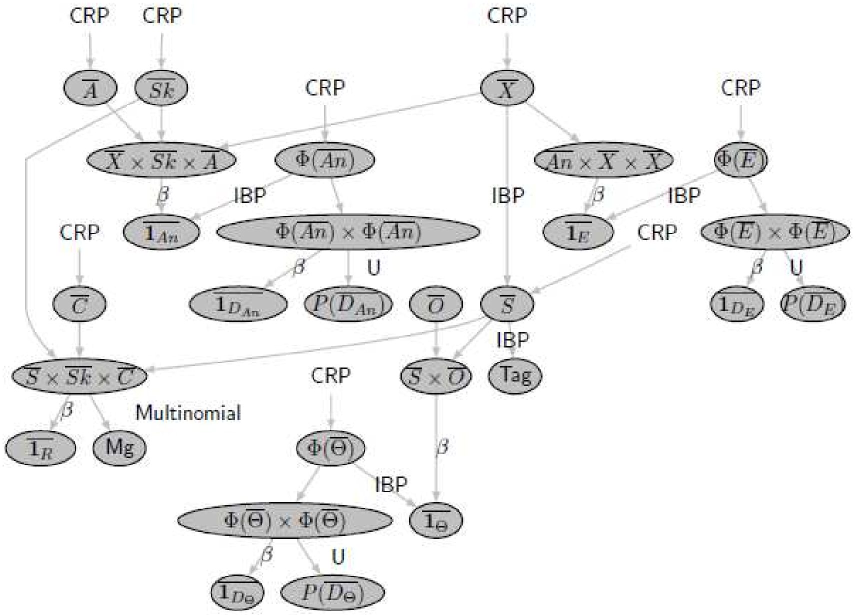


Figure 7: A Non-Parametric Model for Scenario Discovery

First of all, we would hope to discover approximate categories of stakeholders, the kinds of general situational structures they will encounter, the different reward and loss criterion that they have, and the sorts of actions they might take. There could be an infinite number of each of these, but will be found in different percentages in the population with diminishing returns, so let us say they are distributed according to the CRP. The choice of γ for each of these is interesting, as they determine the overall level of exploration, which is discussed in the main text.

$\overline{Sk}|\gamma_{Sk} \sim CRP(\gamma_{Sk})$

$\overline{S}|\gamma_S \sim CRP(\gamma_S)$

$\overline{C}|\gamma_C \sim CRP(\gamma_C)$

$\overline{A}|\gamma_A \sim CRP(\gamma_A)$

Although each structure category may vary in terms of the tags that describe it[22], we expect that some will be used more frequently than others, again with diminishing returns, and thus we can say that the tag distribution is IBP ($tag|\overline{S} \sim IBP(\gamma_{tag,s})$), where $\gamma_{tag,s}$ is small, reflecting the fact that we expect the most salient details to be applied readily and more subtle insights to be rare.

Does a stakeholder experience an impact relative to a concern in a particular stage? We can say that $\overline{\mathbf{1}_R}$ is the distribution over a reward indicator function, specifying that a given criteria is in fact a criteria for a particular stakeholder given a particular structure.

$\overline{\mathbf{1}_R}(\overline{S}, \overline{Sk}, C)|\gamma_{\overline{\mathbf{1}_R}} \sim Beta(\gamma_{\overline{\mathbf{1}_R}}, \gamma_{\overline{\mathbf{1}_R}})$

Here, we can see that $\gamma_{\overline{\mathbf{1}_R}}$ should be very small, say 0.1, as we want it to be very likely or very unlikely that a stakeholder category should have a particular stake given particular conditions, that being the definition of a stakeholder category.

For simplicity, we can treat the magnitude of concern[23] as a multinomial over the discrete categories “strong reward”, “weak reward”, “no reward, no loss”, “weak loss”, and “strong loss”. We will use a uniform prior.

$Mg(\overline{R}(\overline{S}, \overline{Sk}, \overline{C})) \sim Multinomial(0.2, 0.2, 0.2, 0.2, 0.2)$

We can say that a state-of-affairs consists of all of the structures that are currently the case within it. We expect to receive different lists of structures each time we elicit, but again that the variation of structures we find has diminishing returns. Therefore, states-of-affairs could reasonably be represented as a IBP mixture model from a CRP mixture model.

$\overline{X}|\gamma_x \sim CRP(\gamma_x)$

[22] We know that structures contain more, well, structure, than is captured by a tag set. We will save a model that probabilistically builds predicates for future work.

[23] You may wonder why we would represent “no reward, no loss” as a magnitude of concern, and the answer is that it is a matter of relative comparison to the conditions connected by events; consider the phrase “stop hitting me” for example.

$\overline{S}|\overline{X}, \gamma_{x,s} \sim IBP(\gamma_{x,s})$

How should γ_x and $\gamma_{x,s}$ be chosen? If set smaller, it assumes that structures are more likely to be concurrently the case, while if larger, it presumes that different structures are more likely to correspond to different states of affairs. As it stands, we expect both that the states-of-affairs are highly-overlapped, but also that when the stakeholder is describing different outcomes, these descriptions are largely similar except for key factors[24]

The multiplicity of structures in states-of-affairs is not the only time we will want to talk about combinations of elements occurring together. For example, we might want to say some set of events is dependent upon another set of events. For this reason, let us also refer to categories within states-of-affairs as categories of structural combinations ($\overline{X}$ as $\Phi(\overline{S})$). So, when we say $\Phi(\overline{E})$, we mean that sets of events categories occur together in such a way that they can be sampled as an IBP mixture model with a CRP prior, with exploration constants $\gamma_{\Phi E,E}$ and $\gamma_{\Phi E}$, respectively.

Given that we can talk about states-of-affairs, we would like to be able to talk about which actions are possible and relevant for stakeholders in those states-of-affairs, no matter how likely they are. The anticipation indicator function indicates whether an action is anticipated, no matter how likely it is.

$\overline{\mathbf{1}_{An}}(\overline{X}, \overline{Sk}, a)|\gamma_{\overline{\mathbf{1}_{An}}} \sim Beta(\gamma_{\overline{\mathbf{1}_{An}}}, \gamma_{\overline{\mathbf{1}_{An}}})$

If a given action is possible, how likely is to to be undertaken? We would say that it is likely that if the observer knows how to specify it as a possibility, they also likely have a suspicion of whether or not it will be undertaken. For that reason, we can say that the distribution is almost uniform, but not quite, and thus $\gamma_{P(\overline{An})})$ is less than, but close to, one.

$P(\overline{An})|\overline{\mathbf{1}_{An}}(\overline{X}, \overline{Sk}, a), \gamma_{P(\overline{An})} \sim Beta(\gamma_{P(\overline{An})}, \gamma_{P(\overline{An})})$

[24] Indeed, this similarity between possible worlds is a popular philosophical framework for counterfactuals [Lewis, 1973], although it is prior to the intervention-based conception used here.

Of course, as established before, actions do not stand alone, but they are anticipated to have their compliments and substitutes, so there might be a dependence between sets of anticipations, where were need to specify both a likelihood of the dependence (or the indicator) and the likelihood given the dependence.

$\overline{\mathbf{1}_{D_{An}}}(\Phi(\overline{An}),\Phi(\overline{An}))|\gamma_{\overline{\mathbf{1}_{D_{An}}}} \sim Beta(\gamma_{\overline{\mathbf{1}_{D_{An}}}},\gamma_{\overline{\mathbf{1}_{D_{An}}}})$

$P(\overline{D_{An}}(\Phi(\overline{An}),\Phi(\overline{An}))) \sim Uniform(0,1)$

Given a categorical state-of-affairs and some combination of actions, we can ask if another state-of-affairs is the result of that event, and if so, how likely that is:

$\overline{\mathbf{1}_E}(\Phi(\overline{An}),\overline{X},\overline{X})|\gamma_{\overline{\mathbf{1}_E}} \sim Beta(\gamma_{\overline{\mathbf{1}_E}},\gamma_{\overline{\mathbf{1}_E}})$

$P(\overline{D}(\Phi(\overline{An}),\overline{X},\overline{X})) \sim Uniform(0,1)$

Event categories may also have dependencies with both presence and their likelihood, as described before.

$\overline{\mathbf{1}_{D_E}}(\Phi(\overline{E}),\Phi(\overline{E}))|\gamma_{\overline{\mathbf{1}_D}} \sim Beta(\gamma_{\overline{\mathbf{1}_D}},\gamma_{\overline{\mathbf{1}_D}})$

$P(\overline{D_E})(\Phi(\overline{E}),\Phi(\overline{E})) \sim Uniform(0,1)$

The distribution of sensings given observations and underlying structures is similar to anticipations and events, and has similar dependencies.

$\overline{\mathbf{1}_\Theta}(\overline{S},\overline{O})|\gamma_{\overline{\mathbf{1}_\Theta}} \sim Beta(\gamma_{\overline{\mathbf{1}_\Theta}},\gamma_{\overline{\mathbf{1}_\Theta}})$

$P(\overline{\Theta}(\overline{S},\overline{O})) \sim Uniform(0,1)$

$\overline{\mathbf{1}_{D_\Theta}}(\Phi(\overline{\Theta}),\Phi(\overline{\Theta}))|\gamma_{\overline{\mathbf{1}_{D_\Theta}}} \sim Beta(\gamma_{\overline{\mathbf{1}_{D_\Theta}}},\gamma_{\overline{\mathbf{1}_{D_\Theta}}})$

$P(\overline{D_\Theta}(\Phi(\overline{\Theta}),\Phi(\overline{\Theta}))) \sim Uniform(0,1)$

Finally, a category of deferences means that over any set of model elements a deference may be given to a particular category of stakeholders. This deference also has a likelihood, where negative deference is to cast suspicion on the knowledge of a particular stakeholder.

$\overline{\mathbf{1}_{Df}}(\Phi(\overline{M}), \overline{Sk})|\gamma_{\overline{\mathbf{1}_{Df}}} \sim Beta(\gamma_{\overline{\mathbf{1}_{Df}}}, \gamma_{\overline{\mathbf{1}_{Df}}})$

$P(\overline{Df})(\Phi(\overline{M}), \Phi(\overline{Sk})) \sim Uniform(0, 1)$

G: Converting Contingency Paths to Probability Scores

This appendix provides the formalism to support the section "The Relation between Classical and Causal Bayesian Norms" and presumes a strong familiarity with that section. Given this, let us get started with the formalization. Suppose that we are trying to predict some aspect, attribute, or factor, f, which can take one of a discrete number of possible conditions $f_1, f_2, \ldots, f_n$ at a given time t. We can say that the states of affairs in which a given f_i holds is $X_{f_i} \subset X$, and that $X_{f_1}, \ldots, X_{f_n}$ form a partition over X. Further, let us say that there is a set of structural expressions where each structural expression indicates exclusive membership in each of these factor sets, $se_{f,i} \in Se_f \subset Se$. Therefore, the prediction of an outcome is that a particular expression will be true at time t, or in other words $se_{f,i}$ holds for $x[t]$, and the likelihood given in a prediction is $p(se_{f,i}(x[t]))$. Rescober and Tetlock's analysis often focuses on cases that can be resolved from single structures. Given that $S_{f,i}$ is the set of structures that occur only in $X_{f,i}$, then this simple case implies $S_{f,i}$ has at least one member for all $i \in n$, and we can say that the prediction is in reference to that structure, or $p(se_{f,i}(s[t]))$.

When we are establishing the dependencies that would determine a point prediction, we are talking about the set of events that would bring about that point, or $PostSet(se_{f,i})$. We can call a point prediction a likelihood given no event, or the null event prediction, where we say that $e_\emptyset$ is this null event, and that $e_\emptyset \in PostSet(se_{f,i})$ by convention. We also need to account for all dependencies that would boost or sever the event's determination of that point, or $DepSet(PostSet(se_{f,i}))$. Let us say that each phenomena here is weighted according to how strong the events that that produce it are, and how weak the dependences could sever it.

In this scheme, the final point prediction is its normalized weight compared to competing options.

$$p(se_{f,i}(s[t])) = \frac{w_{se_i(s[t])}}{\sum_{i=1}^{n} w_{se_i(s[t])}}$$

The weight normalized weight of the point prediction is determined by the weighted average[25] of the probabilities that could have produced it, including the null event, moderated by the weights of those events. Each single event is the likelihood of that event times the likelihood of its preconditions.

$$w_{se_i(s[t])} = \frac{\sum_{e}^{PostSet(se_{f,i})} w_e p(e) p(Pre(e))}{\sum_{e}^{PostSet(se_{f,i})} w_e p(Pre(e))}$$

The likelihood of the preconditions is the sum over the likely durations of the vent times the likelihood that the preconditions were true at the beginning of that duration.

$$p(Pre(e)) = \sum_{t_\delta=1}^{\infty} \Big(p_{e,t}(t_\delta) \prod_{se_e}^{Pre(e)} p(se_e(s[t - t_\delta])) \Big)$$

The likelihood of each of the precondition expressions $(p(se_e(s[t - t_\delta])))$ is determined recursively, treated as a point prediction at least until $t - t_\delta \leq 0$, at which point the event is in the present or past. Even then the expression may be unknown. For example, it may be a political decision made in secret, or the quantity of undiscovered reserves of a particular resource. In this way of thinking, agreeing that a particular condition did nor did not occur is making an almost certain prediction about the present or past.

The weight of each event is determined the minimum value permitted by the dependencies that determine it (in other words, we choose the maximum severability among those specified). If a combination of events have a stronger severability, then it is appropriate

[25] The weighted average is used because each path is an independent prediction. If each path was not an independent prediction, but instead were predictions of independent causes, there would be a dependency between causes, where the dependency would be $1 - p(c_1)p(c_2)$, or the joint probability.

to introduce a joint event as the antecedent of that dependency. The severability of each dependency is simply the likelihood given the dependency times the inverse likelihood of the antecedents.

$$w(e) = min_{d \in DepSet(e)} \Big(p(d)(1 - max_{e_d \in Ante(d)} p(e_d) w(e_d)) \Big)$$

As a practical matter, we can see that cycles of dependencies can lead to oscillations in this calculation. Therefore, these weights can be calculated iteratively with a dampening constant η, where $0 < \eta < 1$, stopping the iteration when the change of weight reaches some $\epsilon \ll \eta$, yielding the following expression for the sequence of weight changes.

$$w(e)[t] = w(e)[t-1] + \eta(w[e] - w(e)[t-1])$$