- 1 Understanding Scientists' Computational Modeling Decisions about Climate Risk**
- 2 Management Strategies using Values-Informed Mental Models**

ABSTRACT

When developing computational models to analyze the tradeoffs between climate risk management strategies (i.e., mitigation, adaptation, or geoengineering), scientists make explicit and implicit decisions that are influenced by their beliefs, values and preferences. Model descriptions typically include only the explicit decisions and are silent on value judgments that may explain these decisions. Eliciting scientists' mental models, a systematic approach to determining how they think about climate risk management, can help to gain a clearer understanding of their modeling decisions. In order to identify and represent the role of values, beliefs and preferences on decisions, we used an augmented mental models research approach, namely *values-informed* mental models (ViMM). We conducted and qualitatively analyzed interviews with eleven climate risk management scientists. Our results suggest that these scientists use a similar decision framework to each other to think about modeling climate risk management tradeoffs, including eight specific decisions ranging from defining the model objectives to evaluating the model's results. The influence of values on these decisions varied between our scientists and between the specific decisions. For instance, scientists invoked ethical values (e.g., concerns about human welfare) when defining objectives, but epistemic values (e.g., concerns about model consistency) were more influential when evaluating model results. ViMM can (i) enable insights that can inform the design of new computational models and (ii) make value judgments explicit and more inclusive of relevant values. This transparency can help model users to better discern the relevance of model results to their own decision framing and concerns.

Keywords

26 Values-informed mental models, climate change, risk management, decision making under
27 uncertainty
28
29
30

1. INTRODUCTION

Managing climate change risks pose nontrivial decision problems. A number of management strategies exist, but each has advantages and limitations, which impose both temporal and spatial tradeoffs between the potential objectives of stakeholders (e.g., Goes, 2011 and Robock 2008). Furthermore, climate/atmospheric behavior and climate management strategy impacts are rife with deep uncertainty (Keller et al., 2008, Vellinga et al., 2009 and Weitzman, 2011) where decision makers disagree about the appropriate problem framing and how to characterize strategy tradeoffs (McInerney et al., 2012 and Swart et al. 2009). This framing and characterization, though, is imperative for informing decisions about the design and implementation of strategies.

Climate risk management (CRM) strategies include variations of proactive and reactive plans including mitigation (e.g., increased deployment of renewable energy) adaptation (e.g., efforts to prevent flooding in coastal areas) and atmospheric geoengineering (i.e., injecting particles (or particle precursors) into the atmosphere to cool the Earth's surface (Barrett, 2008)). Developing a decision framework for navigating the tradeoffs between such strategies may inform the design of computational models (referred to as models for the remainder of the paper) and decision tools that can analyze and clearly present these tradeoffs. Current analyses of CRM strategies include the use of Earth system models (Sherwood et al., 2014 and Sriver et al., 2012), integrated assessment models (IAMs) (e.g., McInerney et al. 2012 and Nordhaus, 2008), and robust decision making (RDM) analyses (e.g., Hall et al., 2012 and Lempert et al, 2012). These models may take on many different designs and configurations depending on the assumptions made, for instance, about the processes determining the sign and magnitudes of climate feedbacks, economic and social factors that influence greenhouse gas emissions, the atmospheric

chemistry and biogeochemical cycles controlling those emissions, and the impact of those emissions on the climate and, in turn, on humans. When developing such models, scientists (explicitly or implicitly) make decisions about how to characterize the relationships between these assumptions, in particular, how to represent uncertainty in those assumptions and the outcomes (i.e., output) of the model. These decisions are influenced by a number of factors including scientists' preferences, past experience and knowledge.

Understanding the decisions that drive the model designs is arguably important for the informed use of these tools. Typically, important aspects of these decisions are specified in the model description and documentation (see, for example, Nordhaus (2008), Lempert et al. (2012), McInerney et al. (2012), or Sriver et al. (2012)). These descriptions often focus on the explicit decisions and are often silent on arguably important implicit decisions. For example, model documentation usually discusses key model assumptions, but provide, at best, only a very brief discussion regarding *why* the scientist(s) chose a specific metric such as the expected value of the discounted utility of consumption or a specific reliability of achieving an objective (e.g., McInerney and Keller, 2008 and McInerney et al., 2012). Characterizing scientists' "mental models," or how they think about the problem, has hence the potential to improve the understanding of both their explicit and implicit decisions.

A mental model is an internal representation of external reality that is thought to influence reasoning, cognition and decision making (Gentner and Stevens, 1983 and Johnson-Laird, 1983). An individual's mental model includes their experiences, assumptions, beliefs, and biases about the world. In decision making, mental models include an individual's perception of a situation, decision contexts, features in the system, potential solutions and biases (Chermack, 2003). Mental models of a decision situation can be depicted visually as an influence diagram

77 where decisions, and factors that influence those decisions, are shown as nodes, with arrows
78 denoting the direction of influence between nodes (Morgan et al., 2002). This depiction allows
79 for the systematic comparison between individuals' mental models, highlighting how people's
80 thoughts about a decision (e.g., regarding climate risk management strategies) are similar or
81 different.

82 One method for characterizing and comparing individuals' mental models is the Mental
83 Models Approach (Morgan et al., 2002). While originally used to aid in the design of risk
84 communication materials for the general public (e.g., Bruine de Bruin et al, 2009, Downs et al.
85 2004, Fleishman et al. 2010, and Olson et al., 2011), the method has also been used to inform the
86 development of computational models and decision tools for a variety of topics including low-
87 carbon electricity (Mayer et al., 2014), nanoparticles (Morgan, 2005), infrastructure resilience
88 (McDaniels et al., 2008), environmental risk management (Bridges et al., 2013) and flood risk
89 management (Wood et al., 2012). In these applications, the factors that influence the decision
90 being modeled (e.g., which low-carbon electricity or flood risk management strategy to
91 implement) primarily include those related to attributes of the decision alternatives (e.g., costs
92 and benefits of a strategy) or characteristics of the systems that may influence those attributes
93 (e.g., human behavior, climate dynamics).

94 Scientists' values, or their disposition to form a normative attitude (Brennan et al., 2013),
95 toward the factors that influence the decision being modeled are often ignored in the mental
96 models used to inform computational models and decision tools. Values are traditionally treated
97 as exogenous in the Mental Models Approach, essentially being "plugged in" to the factors and
98 decisions after the fact (Morgan et al., 2002). However, individuals do not apply values
99 uniformly to their decisions (e.g., Kahneman et al., 1998 and Kahneman and Tversky 2000). In

fact, Bessette et al. (2016) showed that individuals' values related to CRM are applied differently depending on the strategy being evaluated. When values are actually considered within mental models, they usually only include a small subset of ethical values such as "priorities to protect public health/welfare... [and] environmentally sensitive areas" (Wood et al., 2012, p. 1357). Values can range, though, from well-articulated normative commitments to dispositional states that are unacknowledged by individuals (Diekmann and Zwart, 2014). For example, values related to CRM strategies for New Orleans have been shown to range from 'basic needs and survival' to 'sense of place' to 'stakeholder engagement' (Bessette et al., 2016).

Values play a central role in the development of computational models (Diekmann and Zwart, 2014, Fleischmann and Wallace, 2006 and Wallace, 1994). As an illustration, a scientist who strongly values future generations (i.e., intergenerational justice) may develop a model that includes outcomes (e.g., costs and benefits) specific to future people, and long timescales that can accommodate such outcomes. This need for long time scales then can motivate the scientist to value models that are more parsimonious and easier to calibrate, compared to more complex models that include more processes, but have to rely more on expert judgments and are more difficult to calibrate (Oppenheimer et al., 2008). As a second example, consider the decision about the level of model complexity and resolution that climate modelers often face. Increasing the model complexity and/or resolution can enable an improved representation of processes but typically comes at the cost of a decreased number of feasible model evaluations. These value judgments also influence other factors scientists consider. As in the first example, values may influence *why* scientists include certain attributes of the decision alternatives (e.g., costs and benefits of a strategy) and characteristics of the systems (e.g., human behaviors) in their models. Users of decision support tools may not explicitly realize these value judgments.

Understanding *why* scientists think about the tradeoffs (e.g., between CRM strategies) the way they do may provide a window into the value judgments of the scientists who develop computational models and decision support tools. Such insights can help to inform the design of new models and tools that make these value judgments explicit. This transparency can allow decision makers who use these models and tools to better discern the relevance of model results to their own decision framing and concerns.

In this study, we use an augmented mental models research approach named values-informed mental models (ViMM) (Bessette et al., 2016) to capture how a set of scientists thinks about modeling CRM strategy tradeoffs and *why* – that is, how their values influence this thinking. ViMM includes an explicit focus on values in order to illustrate how they influence an individual's representation of the decision situation. Rather than assuming that values are only relevant once the problem has been identified and objectives have been clarified, ViMM recognizes that an individual's representation of a situation is interdependent with their values.

We conducted semi-structured interviews with eleven interdisciplinary scientists (e.g., climate scientists, economists, decision analysts) who lead projects under the NSF-sponsored research network for Sustainable Climate Risk Management (SCRIM) (see www.scrimhub.org for more details) and have published considerably in areas relevant to climate risk management. These interviews were transcribed and qualitatively analyzed to isolate the scientists' modeling decisions. The analysis sheds light on the sort of factors considered by scientists, including how those factors influence their decisions. These factors include those found in more traditional mental models applications, such as strategy attributes and systems (e.g., social or physical systems), but also those related to scientists' ethical and epistemic values. As more transdisciplinary networks are being formed to undertake (and model) high-impact issues, such

as the SCRIM network’s work on the issue of CRM discussed here, explicitly identifying the values and other factors driving their modeling decisions will become increasingly important for the consumers of those models – that is, the members of the many stakeholder groups who will be impacted by such model results.

2. METHODS

2.1 Interviews

We conducted one-on-one interviews with eleven SCRiM scientists over the phone, which lasted between one and two hours. To restrict the scope of the discussion, interviews followed a semi-structured protocol (see Supplementary Material) that was designed to highlight the critical tradeoffs between CRM strategies including their costs, benefits, limitations, uncertainties and relevant values. Semi-structured interview protocols balance the standardization of a structured elicitation with the flexibility of an open-ended response mode. The protocol used here included twelve questions that were ordered in progressive specificity. Initial questions were more general, not specifically focused on computational modeling decisions, and included probes on strategy outcomes, uncertainties, assumptions and their related values (e.g., “How would you compare and contrast the three management strategies?”). Later questions focused on specific modeling areas of interest including performance measures, decision tool design, and model users (e.g., “What performance measures are available and could or should be used to compare/contrast these management strategies?”). Interviewees were provided the semi-structured protocol prior to the interview. Interviews were audio-recorded and transcribed.

Because the set of interviewees had diverse expertise, subsequent to the interview, scientists were asked to self-assess their level of expertise in seven areas (ethical and epistemic analysis, technology assessment, uncertainty quantification, earth systems analysis, integrated assessment, advanced computational methods, and stakeholders and decision makers) on a scale from 1 (low expertise) to 5 (high expertise). Additionally, they provided their years of experience within their field of expertise.

2.2 Participants

As shown in Table 1, on average, interviewees (n=11) described themselves as having the most expertise in uncertainty quantification (Mean=3.9) and the least expertise in technology assessment (Mean=2.5) as well as ethical and epistemic analysis (Mean=2.5). At least one interviewee indicated they had high expertise (=5) in each area except for technology assessment. Interviewees had a diverse level of experience working in their area of expertise, with three having between 1-10 years, five having 11-20 years and three having more than 21 years of experience.

Table 1. Interviewee Expertise

	Mean Rating $\pm$ SD	Rating (Min, Max)
Uncertainty quantification	4.1 $\pm$ 1.2	(2, 5)
Advanced computational metrics	3.5 $\pm$ 1.6	(1, 5)
Integrated assessment	3.3 $\pm$ 1.3	(1, 5)
Earth systems analysis	3.2 $\pm$ 1.5	(1, 5)
Stakeholders and decision makers	2.5 $\pm$ 1.3	(1, 5)
Coupled ethical-epistemic analysis	2.5 $\pm$ 1.1	(1, 5)
Technology assessment	2.5 $\pm$ 0.9	(1, 4)

Note: Scientists self-assessments of their level of expertise were provided on a scale from 1 (low expertise) to 5 (high expertise)

2.3 Qualitative Analysis

We conducted a formal content analysis of the interview transcripts to identify a set of topic areas and themes. We first used a structural coding approach to assemble an a priori list of topic areas based on the structure of the interview protocol profiles.⁽³⁶⁾ Subsequently, we refined the coding scheme based on reviewing the interview transcripts, following a grounded theory approach.⁽³⁷⁾ Our efforts resulted in three sets of codes that would encompass those in a traditional mental model:

- eight modeling decisions (e.g., define objectives/outcomes of interest, define scope, interpretation and evaluation judgments),
- three systems (i.e., geopolitical dynamics and human behavior, economic/market dynamics, climate/ecological dynamics), and
- six strategy attributes (e.g., effectiveness/feasibility, costs, impacts).

In addition, we defined two sets of values codes:

- nine ethical values (e.g., fairness, welfare, intergenerational justice), and
- nine epistemic values (e.g., robustness, reliability, consistency).

A full list of these codes and their definitions are given in Supplementary Material.

We developed a consensus approach to coding, including application of particular codes to sections of text from the interview transcripts. Subsequently, each transcript was coded for the decisions, systems and strategy attributes by one coder, who iteratively refined the application of codes based on review by a second coder. Two different coders with expertise in ethical and epistemic values then independently applied the ethical values and epistemic values codes. These two coders double coded two exemplary transcripts across the full set of codes as

an inter-rater reliability exercise and yielded a Fleiss' kappa measure of 0.40, indicating moderate agreement between reviewers (Landis and Koch, 1977).

Transcript text was segmented into quotes, with each quote representing a different concept discussed by interviewees. Codes were then applied to each quote in the text only if the coded concept was prevalent. Multiple codes could be applied to each quote. We treated each of the 36 codes as a potential node in the mental model influence diagram. To determine where influences exist between the decision nodes and the nodes representing factors that influence decisions, we isolated text from the transcripts in which two codes were applied to the same text. The sections of text were reviewed to determine the existence and direction of influences. We then constructed the mental model using the nodes and arrows of influence.

3. Results

3.1 Scientists' Modeling Decisions

We found that scientists use a similar decision framework to each other to think about modeling climate risk management tradeoffs, including eight specific decisions that are discussed below and detailed in Table 2. Figure 1 provides a general depiction of how scientists talked about modeling decisions, presented using the more traditional mental models methodology (Morgan et al., 2002). The decisions highlighted in the figure note those nested mental models explored in later sections. Scientists tended to discuss eight different specific modeling decisions related to characterizing the tradeoffs of the CRM strategies. We characterized these specific modeling decisions into two more general decision stages that may be iterative in nature: model structuring (i.e., decisions related to designing the model) and model evaluation (i.e. decisions related to using the modeling results). Five of the specific

modeling decisions (i.e., define model objectives/outcomes of interest, define model strategies, define model scope, define model relationships and define model inputs) encompassed the model-structuring phase. For instance, one scientist discussed the decision of defining model objectives (coded as define objectives/outcomes of interest):

“...this is painting a broad brush but a lot of attempts at optimized strategies or policy are based on aggregate economic measures that...don't do a particularly good job of reflecting human wellbeing..”

Table 2. Eight Modeling Decisions used by CRM Scientists

Decision	Decision Stage	Description
Define scope	Model Structuring	Decision-making timescale and spatial parameters and boundaries
Define strategies	Model Structuring	Strategy design and implementation characteristics
Define objectives/ outcomes of interest	Model Structuring	Criteria and goals of strategy implementation and outcomes (e.g., wellbeing, biodiversity, costs, efficiency, certainty)
Define relationships	Model Structuring	Discussion of model design characteristics and model design omissions, including functions and equations
Define inputs	Model Structuring	Discussion of model input characteristics and model input omissions, not covered by scope (e.g., uncertainty characterization)
Interpretation and evaluation judgments	Model Evaluation	Interpretation and evaluation of model inputs and model results (e.g., trusting model results)
Recommendations	Model Evaluation	Experts' strategy preferences based on model recommendations
Communication	Model Evaluation	Related to what and how the expert communicates model recommendations and results

Figure 1. Scientists' Mental Models of Computational Modeling Decision for CRM Strategies

Note: Decisions are represented as rectangles, processes as diamonds, with arrows representing direction of influence. Highlighted decisions emphasize those nested mental models explored in detail in Figures 3 and 4

Another scientist discussed implementing a systematic approach to scoping of the model (coded as define scope):

“...And so one principle that I think is important is...we have to really think about it in a systematic way, which means that it has to be an integrated analysis. So if you do an analysis that is very strong on the technical point of economic analysis or scientific analysis, but is weak on the communications side or on the feedback, it is perfectly fine and would follow principles and values in a certain discipline, but for the problem at hand, it might actually not be very helpful.”

While we found evidence of each of the specific decisions in the model-structuring phase influencing each other, scientists generally invoked the model structuring stage by first defining

the model's scope. The scope influences how the strategies are defined followed by an identification of the objectives of the model and outcomes of interest. Next, scientists combine all of this information to define the modeling relationships and as a final step, the inputs needed for the model. We add an arrow from the defining inputs to defining scope phase to signify that this process can be iterative in nature and that as scientists make these successive decisions they may reenter the model structuring stage to revise earlier decisions based on later structuring phases. Based on this model structuring, scientists could then run computational models that evaluate the tradeoffs between CRM strategies.

After running the model, scientists enter the model evaluation phase. In this phase, scientists interpret model results and make judgments about the evaluation (coded as interpretation and evaluation judgments), which are shaped by decisions made during the model-structuring phase. For example, one scientist discussed how s/he judges a specific model:

"Now, this [model] ...excludes contributions from fast changes of ice sheets. And those are known things that are not well represented currently in the model...if people take those [models] at face value, they neglect known unknowns. And the known unknowns typically tend to widen the range of possible outcomes...which means potential error risks are larger, and people might have wrong risk assessments."

These judgments lead scientists to make recommendations about, for example, the implementation of certain strategies or modification to the model (coded as recommendations). One scientist provided recommendations related to mitigation:

287 “The way I look at it is we know what the solution is: it’s mitigation. And if we look at costs, on a
288 planetary basis it’s going to save us a lot of money but certain selfish interests are going to lose
289 money in the short term and so they’re the ones that are standing in the way of doing anything
290 about it.”

291
292 Scientists also discussed communicating about their judgments, recommendations and model
293 structuring (coded as communication).

294
295 “And I think that it depends on how you’re talking about mitigation and what decision makers
296 are trying to seek—whether it’s a general population messaging or you’re talking about specific
297 energy conversion to renewables coordination across sectors...”

298
299 The recommendations and communications may lead to revisions in scientists’ model
300 structuring, producing an iterative process. Further, every one of the eight scientist decisions
301 depicted in this influence diagram (Figure 1) is influenced by a set of values, strategy attributes
302 (e.g., cost) and systems (e.g., social or physical systems).

304 **3.2 Factors Influencing Scientists’ Modeling Decisions**

305 *3.2.1 Scientists’ Consideration of Values*

306 The ViMM approach enabled us to identify nine ethical and nine epistemic values during
307 the interviews (defined in full in Supplementary Material). For each of these 18 values, we
308 calculated their quotation *prevalence* within each transcript, or the percent of quotes within a
309 given transcript that included a code for that value. For instance, if one transcript was segmented

into 100 quotes and included 10 quotes related to the ethical value of welfare, its *prevalence* would be 10%. Additionally, for each of the 18 values, we calculated their *coverage*, or the number of scientists (out of 11) who mentioned that value at least once during their interview. Figure 2 presents the values' *coverage* and mean, minimum and maximum *prevalence* for the 11 scientists.

Overall, epistemic values (i.e., values relevant to acquiring knowledge) were discussed by scientists in lesser prevalence than ethical values (i.e., values related to how one should act). While every scientist discussed at least one ethical and one epistemic value during their interview (i.e., *coverage*=11) for both ethical and epistemic values overall), on average, 16.8% of scientists' quotes included an ethical value, while only 9.0% of their quotes included an epistemic value (i.e., mean prevalence for each). However, this prevalence did vary by scientist for types of values. One scientist discussed some type of ethical value in 43.8% of her/his quotes (the maximum), while another only discussed ethical values in 6.7% of her/his quotes (the minimum).

For ethical values, welfare (i.e., concerns about effects of strategies on enjoyment/possession/promotion of important goods) and intergenerational justice (i.e., concerns about tradeoffs between the distribution of impacts on current vs. future generations) were mentioned by all scientists (*coverage*=11), with distributive justice (i.e., concerns about distribution of costs and benefits of various strategies on people) mentioned by nearly all scientists (*coverage*=10). Welfare also was the most prevalent value mentioned in the transcripts. However, the prevalence of discussion about welfare varied widely by the scientists, from 1.5% to 12.5%, suggesting that some scientists consider it more in model design and evaluation than others.

Figure 2. Coverage and Mean Prevalence of Value Codes

Note: Coverage is the number of scientists (out of 11) who mentioned that value at least once during their interview. Prevalence is the percent of quotes, within a given transcript, that included a code for that value. In the prevalence graph, the line represents the range of prevalence across scientists, with the solid circle marker, and value directly above, representing the mean. Summary statistics for prevalence only include those transcripts where the value was coded at least once.

While democratic values (i.e., concerns related to the decision-making processes being democratic) were only mentioned by a little over half of the scientists, it was the second most prevalent ethical value, suggesting an “all or nothing” mentality with this value. That is, if scientists were at all concerned about representing democratic values in models, these results suggest they placed a large priority on the value. The precautionary principle (i.e., concern over risks being so serious they are cited as reason not to pursue some strategy) and basic rights protection (i.e., concern that rights to life, liberty, etc., are protected or promoted) were mentioned by few scientists, and were not particularly prevalent in those scientists’ interviews. Examples of ethical values mentioned include:

“But we also need someone who is thinking about those who can’t speak ... someone who is, in a sense, standing up for non-human life forms—you know, animals, ecosystems, etc. and taking those into consideration.” (Welfare)

“We will have serious impact, serious costs if we do not mitigate and we just do adaptation, we will be very able to adapt while some of the poor regions will not be able to adapt, will not have the money and they will suffer much more.” (Distributive justice)

“So that I would say a fair and equitable policy decision has to be made by people who are capable of putting self-interest aside, who are informed by a wide range of value decisions and who have a sense of...the balance of those values amongst the people being affected.”

(Democratic values)

Prediction (i.e., concerns for the model's predictive power) was the most prevalent epistemic value, while consistency (i.e., concerns for models being consistent and not contradictory) was discussed by most scientists. Consistency also contained the largest range of prevalence among scientists, with at least one scientist mentioning it in 10.4% of quotes and some scientists not mentioning it at all. Bias immunity (i.e., concerns about the degree of sensitivity of the model to certain biases) and simplicity (i.e., the belief that simpler theories are better) were mentioned by very few scientists, and were not particularly prevalent in those scientists' interviews. Examples of epistemic values mentioned include:

"The general issue of global coverage of data is critical and particularly for the long-term mitigation and geoengineering issues, in order to say whether a climate model is behaving well, you have to be able to say whether or not the South Pacific or the Southern Ocean or the interactions between the ice sheets or the sea ice margins and the ocean are behaving as the models are predicting." (Prediction)

"And so you want tools which help people understand the challenges they face. You want to do it in such a way that it admits to multiple worldviews and values and interests but in a way that is constrained by the science and the available information but that doesn't present that in an overconfident way." (Consistency)

3.2.2 Influences on a Specific Decision

Each specific modeling decision represented in Figure 1 is influenced by a different combination of values, strategy attributes and systems. Figures 3 and 4 represent that influence

for two specific modeling decisions, chosen (out of the eight total decisions) because they clearly illustrate different patterns of influences on the decisions: define model objectives/outcomes of interest, and model interpretation and evaluation judgments. These patterns mirror those found for the other modeling decisions in the model structuring and model evaluation phases (results not shown), in that defining model objectives/outcomes of interest are more heavily influenced by ethical values, whereas model interpretation and evaluation judgments are more heavily influenced by epistemic values. We calculated the *coverage* of this influence, or the number of scientists (out of 11) who discussed a value, strategy attribute or system as influencing each of these decisions. In the figures, coverage of these influences is represented as a progressive shading of the value, strategy attribute and system nodes, where darker shading signifies a greater number of scientists discussing that factor as influencing their decision.

Of the three systems coded, the transcript content analysis suggests that climate/ecological dynamics had the largest influence on the modeling decision of how to define objectives/outcomes of interest, with nine scientists discussing the concept. Coverage of the other two systems influence the decision of how to define model objective nearly as much, suggesting that while climate/ecological dynamics may be the most influential factor, all systems' influences were well represented by the diverse group of experts. As an example of the influence of climate/ecological dynamics, one scientist discussed temperature change and ecosystem impacts as outcomes of interest:

"The other issue with geoengineering, adaptation and mitigation is it depends on which of the big factors are important. Are you interested in just plain temperature change and the economy?...Are we interested in ecosystem impacts? Are we interested in services associated

with the natural systems?” (Coded as define objectives/outcomes of interest & climate/ecological dynamics)

Figure 3. Values, Strategy Attributes and Systems Influencing Scientists’ Decisions Related to Defining Objectives/Outcomes of Interest

Note: The progressive shading and numbers shown in parentheses represent the coverage of these influences (or the number of scientists out of 11 who discussed that value, strategy attribute or system). Darker shading signifies a greater number of scientists discussing that factor as influencing their decision.

Figure 4. Values, Strategy Attributes and Systems Influencing Scientists' Decisions Related to Interpretation and Evaluation Judgments

Note: The progressive shading and numbers shown in parentheses represent the coverage of these influences (or the number of scientists out of 11 who discussed that value, strategy attribute or system). Darker shading signifies a greater number of scientists discussing that factor as influencing their decision.

Our analysis also suggests that considerations about strategy effectiveness/feasibility exerted the most influence of the strategy attributes (*coverage*=9) on the decision of how to define model objectives/outcome of interest, followed by stakeholders and decision makers and strategy impacts (*coverage*=8) with potential strategy innovation and policies that could facilitate or inhibit policies exerting the least influence (*coverage* =1). The lack of suggested influence of

the latter two strategy attributes may be a result of their lack of coverage overall, as discussed in the previous section.

Similar analyses for the value codes suggest that welfare and distributive justice were equally the two most influential ethical values toward the decision of how to define model objectives (*coverage*=5 for both). For example, one scientist discussed human welfare as an objective when evaluating the CRM strategies:

“In general, for addressing human life and the ability for humans to live on the planet, then you get into the standard sustainability question where the most important outcomes are: are we able to get food, water and shelter, and...how are these things impacted by climate change? And of the three strategies, which ones are most appropriate for dealing with human activities on the planet?” (Coded as define objectives/outcomes of interest & welfare)

Interestingly, scientists did not mention in the interviews a number of epistemic values when they talked about components of the modeling decision. These include simplicity, testability/falsifiability (i.e., concern about whether the model has been rigorously tested), bias immunity and reliability (i.e., concern about whether modeling techniques reliably produce good models) (*coverage*=0 for all four values). Prediction was mentioned by the most scientists as exerting influence on the decision of how to define model objectives (*coverage*=4).

Overall, our analyses suggest that ethical values tended to have a more explicit influence on scientists’ decisions of *how to define objectives and outcomes* of interest than epistemic values. While at least one scientist discussed each ethical value in terms of how to define model objectives, four epistemic values were omitted completely in this discussion.

In contrast, Figure 4 suggests that epistemic values exerted a stronger influence than ethical values on decisions related to *model interpretation and evaluation judgments*. Every epistemic value except simplicity was covered by at least one scientist as an influence on model evaluation judgments, while the ethical values of equity (i.e., concern that there are no worrying disparities in the distribution of costs and benefits), precautionary principle and basic rights protection were not mentioned by any scientist as exerting influence on their model evaluation judgments. Of all values, consistency showed the most influence (*coverage*=8). For example, one scientist described the importance of having consistent knowledge when making decisions about future sea level rise:

“People don’t always agree on values but there are ways to adjudicate between those disagreements...[and get] as robust, empirical understanding as we possibly can get. We need just as much knowledge as we possibly can have. So if we’re making decisions about sea level rise we need to know as much as we possibly can and about the right kind of stuff.” (Coded as interpretation and evaluation judgment & consistency)

For strategy attributes, stakeholders and decision makers were mentioned by some scientists as exerting influence on their model evaluation judgments (*coverage*=8), while strategy impacts had the most coverage when intersected with evaluation judgments (*coverage*=9). These results may suggest that when interpreting model results, scientists are most concerned with the impacts (e.g., costs and benefits) of the strategies and how those impacts affect stakeholders involved. For instance, one scientist described how s/he evaluates integrated assessment models based on the models’ handling of specific strategy impacts.

483
484 *“So in almost all of these models, any climate change mitigation measure is always a*
485 *redistribution from the rich future generations to the poor current generation. So most of us feel*
486 *a little uneasy with this implicit assumption of the integrated assessment models, and I think that*
487 *is something very important to analyze. It is underlying that economic growth will outgrow any*
488 *sort of climate problem, and that’s in most of these Integrated Assessment models what*
489 *happens.”* (Coded as interpretation and evaluation judgment & impacts)

490
491 **4. DISCUSSION**

492 Our results highlight how scientists make a number of (explicit and implicit) decisions
493 when thinking about how to structure and evaluate models. In this specific case, we found that a
494 transdisciplinary sample of scientists, working together on issues related to CRM tradeoffs,
495 consider eight specific modeling decisions, and that these decisions are each influenced by a
496 different set of values, strategy attributes and system considerations. These decisions range from
497 defining the CRM strategies (that would be evaluated by the model) to choosing how and what to
498 communicate about the model and its results. We found that scientists’ progress through a
499 model-structuring phase before they evaluate the models and their results to make choices
500 between CRM strategy tradeoffs, and ultimately the strategies to recommend for implementation.
501 The model-structuring and model-evaluation phases occur in an iterative fashion for our
502 scientists, in which the interpretation of model results flow back into the model-structuring
503 phase.

504 Many of the values, strategy attributes and system considerations vary among the
505 scientists (as demonstrated by often less-than-universal coverage). That said, we found a few

distinct commonalities. For instance, every scientist considered the ethical values of welfare and intergenerational justice when making decisions about modeling CRM tradeoffs, with most also including distributive justice and the epistemic value of consistency in their decision making. These four values, as well as democratic values and prediction were prevalent in approximately 10% or more transcription quotes for at least one scientist, as measured by their maximum *prevalence*. The high *coverage* and *prevalence* of these values suggests that they are of central influence to the decisions made by scientists when designing and evaluating models. For instance, almost half of the scientists implied that concerns about welfare and distributive justice influenced how they decided to define model objectives and more than two-thirds of scientists implied that the consistency of the model was a major consideration when evaluating its results. While every value was mentioned by at least one scientist (a result of creating the values codebook after reading a sample of the interviews), a number were mentioned by fewer than half of the scientists, suggesting that scientists' modeling decisions are influenced by different sets of values.

The variation in scientist's individual values-informed mental models (ViMMs) found in this study is perhaps not surprising. While these scientists are each part of the same research network, the network is by design transdisciplinary and scientists' expertise and experience are considerably diverse. Our results suggest that these scientists use a similar decision framework to think about modeling CRM tradeoffs, but their decisions may vary because of the influence from different beliefs, preferences and values. This may be seen as having both positive and negative implications. While the results suggest that no one may have a universal perspective, it also suggests the value of these very transdisciplinary endeavors, since they help to support a more comprehensive treatment of a very complex issue.

Overall, ethical values were discussed in greater prevalence than epistemic values in the context of modeling decisions, potentially suggesting that ethical values have a more explicitly formative role for these scientists than epistemic values. While further research would be required to support a full interpretation of these results, one hypothesis for explaining them is that ethical values are viewed as more complex or perhaps more controversial than the epistemic aspects of this decision set. Another interpretation could be that these scientists care very deeply about the ethical issues related to climate change risks – and such issues are more at the center of their concerns than the epistemic elements of their mental models – making it not surprising that they discussed the former more than the latter. This dominance, though, is suggested to be decision-specific, as seen in the comparison of which values’ influence the two exemplary decisions we presented in our analysis. Whereas our results suggest that the decision of how to define model objectives is most influenced by ethical values, epistemic values may be more influential towards decisions related to model interpretation and evaluation judgments. Following from our first hypothesis above, these results may imply that the ethical values are viewed as more complex or controversial for defining model objective, while the epistemic values take on that role for model interpretation.

Systems and strategy attributes were also factors suggested to influence scientists’ decisions. For the two specific decisions we present, for instance, scientists implied that the impacts (e.g., costs and benefits) of the strategies and how those impacts affect stakeholders and decision makers were of great importance. *Coverage* of these two strategy attributes is greater than those for any value except for considerations related to consistency for the model interpretation decision. This first suggests that scientists may be giving equal (or in many cases greater) priority to modeling choices related to its consistency and to its strategy attributes.

However, for the other values with less *coverage*, the result may suggest that scientists are not explicitly considering values (as much as the systems and strategy attributes) in their model design and evaluation (and thus, they did not discuss it as much during their interviews). The latter conclusion is consistent with current model descriptions and documentation available (Lempert et al., 2012, McInerney et al., 2012, Nordhaus, 2008, and Sriver et al., 2012), which appear to represent *what* modeling decisions were made but not *why* they were made.

Of course, such results should be interpreted with caution. Our interviews only provide insight into what scientists say, not what they actually do. Many studies have shown that a respondent's intention to behave a certain way does not always result in the occurrence of that behavior (Ajken, 2011 and Fishbein, 1979). Indeed, the only method of confirming a behavior, such as that of making a modeling decision, is through direct participant observation (becker and Geer, 1977). Participant observation has also been shown to provide more information and richer descriptions of participants than interviewing alone (Tope et al., 2005). Conversely, observation alone would only provide limited insights into scientists' internal mental models, which are inherently conceptions that may or may not be reflected in behavior.

In this study, the explicit representation of the ViMMs of an interdisciplinary network (SCRiM) of scientists formed to undertake (and model) the high-impact issues of CRM may inform an understanding of *why* they may be making specific decisions when designing a model. For instance, the ViMMs illuminate why they may have modeled relationships in a specific way or chosen a specific representation of uncertainty in model assumptions. This explicit, transparent, and systematic representation of value judgments is something that was largely missing from prior mental models analyses (Bessette et al., 2016). Other large transdisciplinary networks may also benefit from using the ViMM method to explicitly identify the values and

other factors driving their decisions and products they generate (such as models). These products likely reflect the value judgments of the network of scientists, which can now be known to decision makers and consumers who may use the products. Decision makers, who likely will not hold the same values, beliefs and preferences as the scientists, can therefore determine the relevance of these products and whether they are useful for the decisions at hand. Future studies using ViMM to inform model design could evaluate such outcomes of the ViMM method.

The application of ViMM (Bessette et al., 2016) presented in this paper is potentially useful, but, as designed, was unidirectional, providing information only on the value choices of scientists. This knowledge could certainly be useful to decision makers who may use these models, helping them to decide whether the models are applicable to their value choices. However, similar to the iterative, inclusive and prescriptive design of risk communications recommended as best practice (Fischhoff, 1995 and Morgan et al., 2002), ViMM has the potential to be applied in an iterative manner. That is, ViMM could be applied to decision makers, stakeholders and scientists to ensure that models (e.g., about CRM strategy tradeoffs) best account for the beliefs, values and preferences not only of the model builders, but also of the decision makers who ultimately use them and the stakeholders who may be affected by model-driven decisions. In the traditional mental models approach, mental models of the lay audience inform the design of communications and are compared to the experts' understanding of the problem to include all relevant decisions and beliefs in the resulting communications (Morgan et al., 2002). Similarly, ViMMs of decision makers and stakeholders could inform the design of such models, and a comparison of the ViMMs of scientists, decision makers and stakeholders could highlight the range of values and beliefs to be represented in these models.

For example, ViMMs of New Orleans stakeholders about CRM suggest that values related to preserving New Orleans' unique culture, traditions and historically significant neighborhoods are nearly as important to those stakeholders as those related to human welfare (Bessette et al., 2016). While our scientists' ViMMs were not specific to the New Orleans area, a 'sense of place' value was not identified, even though climate change risks that could cause damage to culturally relevant areas were discussed. This gap in our scientists' ViMMs could imply that, without some intervention, these values would not be captured in the CRM models they develop. Comparing such ViMMs to identify value gaps could help refine CRM models such that these values would be included.

By introducing values into the representation of CRM models, ViMM provides a view not only into what decisions are made when designing these models, but *why these decisions are made*. Our results suggest that scientists hold a large variation in values, beliefs and preferences regarding CRM. Once future studies capture the ViMM of decision makers, computational models and decision tools may be designed that are balanced among the many stakeholders, presenting decision- and values-relevant information.

5. CONCLUSION

This study used values-informed mental models (ViMM) (Bessette et al., 2016) to capture how a set of scientists thinks about modeling CRM strategy tradeoffs and how their values influence this thinking. Our qualitative analysis of interviews with eleven CRM scientists suggests that they use a similar decision framework to each other to think about modeling CRM tradeoffs, including eight specific decisions ranging from defining the model objectives to evaluating the model's results. Importantly, the ViMM method highlighted the ethical and

epistemic values that appear to influence scientists' model decision making. Our results suggest that ethical values tended to be more influential on decisions related to model structuring, while model evaluation decisions were shaped more by epistemic values. The ethical values related to welfare and intergenerational justice were discussed by all scientists in the context of CRM modeling. These values, in addition to democratic values and prediction, were also discussed the most often during our interviews.

In addition to highlighting the types of values that scientists' may invoke as they develop CRM models, this application of the ViMM method suggests that, more generally, it can make value judgments more explicit and enable the use of those insights to inform the design of new computational models.

631 **REFERENCES**

- 632 Ajzen I., 2011. Theory of planned behavior. *Handb. Theor. Soc. Psychol.* 1, 438-459.
- 633 Barrett, S., 2008. The incredible economics of geoengineering. *Environmental and Resource*
634 *Economics* 39 (1), 45-54.
- 635 Becker, H., Geer, B., 1977. Participant observation and interviewing: A comparison. *Human*
636 *organization* 16 (3), 28-32.
- 637 Bessette, D., Mayer, L.A., Cwik, B., Vezer, M., Keller, K., Lempert, R., Tuana, N., 2016.
638 Building a values-informed mental model for New Orleans Climate Risk Management. *Risk*
639 *Analysis*, In press.
- 640 Brennan, G., Eriksson, L., Goodin, R.E., Southwood, N., 2013. Explaining norms. Oxford
641 University Press, Oxford.
- 642 Bridges, T.S., Kovacs, D., Wood, M.D., Baker, K., Butte, G., Thorne, S., Linkov, I., 2013.
643 Climate change risk management: A mental modeling application. *Environment Systems and*
644 *Decisions* 33 (3), 376-390.
- 645 Bruine de Bruin, W., Güvenç, Ü., Fischhoff, B., Armstrong, C.M., Caruso, D., 2009.
646 Communicating about xenotransplantation: Models and scenarios. *Risk Analysis* 29 (8),
647 1105-1115.
- 648 Chermack, T. J. Mental models in decision making and implications for human resource
649 development. *Advances in Developing Human Resources* 5 (4), 408-422.
- 650 Diekmann, S., Zwart, S.D., 2014. Modeling for fairness: A rawlsian approach. *Studies in History*
651 *and Philosophy of Science Part A* 46, 46-53.

652 Downs, J.S., Murray, P.J., Bruine de Bruin, W., Penrose, J., Palmgren, C., Fischhoff, B., 2004.
653 Interactive video behavioral intervention to reduce adolescent females' std risk: A
654 randomized controlled trial. *Social Science & Medicine* 59 (8), 1561-1572.

655 Fischhoff, B., 1995. Risk perception and communication unplugged: Twenty years of process.
656 *Risk Analysis* 15 (2), 137-145.

657 Fishbein, M., 1979. Theory of reasoned action: some applications and implications. In Howe H.,
658 Page, M. (Eds.), *Nebraska Symposium on Motivation*, University of Nebraska Press,
659 Lincoln, pp. 65-116.

660 Fleischmann, K.R., Wallace, W.A., 2006. Ethical implications of values embedded in
661 computational models: An exploratory study. *Proceedings of the American Society for*
662 *Information Science and Technology* 43 (1), 1-16.

663 Fleishman, L.A., Bruine de Bruin, W., Morgan, M.G., 2010. Informed public preferences for
664 electricity portfolios with ccs and other low-carbon technologies. *Risk Analysis* 30 (9), 1399-
665 1410.

666 Gentner, D., Stevens, A.L., 1983. *Mental models*. Lawrence Earlbaum Associates Inc., New
667 York.

668 Glaser, B. G., Strauss, A.L., 2009. *The discovery of grounded theory: Strategies for qualitative*
669 *research*. Transaction Publishers, Piscataway, NJ.

670 Goes, M., Tuana, N., Keller K., 2011. The economics (or lack thereof) of aerosol
671 geoengineering. *Climatic change* 109 (3-4), 719-744.

672 Hall, J.W., Lempert, R.J., Keller, K., Hackbarth, A., Mijere, C., McInerney, D.J., 2012. Robust
673 climate policies under uncertainty: A comparison of robust decision making and info gap
674 methods. *Risk Analysis* 32 (10), 1657-1672.

675 Johnson-Laird, P.N., 1983. Mental models: Towards a cognitive science of language, inference,
676 and consciousness. Harvard University Press, Cambridge, MA.

677 Kahneman, D., Shkade, D., Sunstein, C.R., 1998. Shared outrage and erratic awards: the
678 psychology of punitive damages. *Journal of Risk and Uncertainty* 16, 49-86.

679 Kahneman, D., Tversky, A., 2000. Prospect theory: an analysis of decision under risk. In
680 Choices, Values, and Frames, Kahneman, D., Tversky, A., (Eds.), Cambridge University
681 Press, New York, pp. 17-43.

682 Keller, K., Yohe, G., Schlesinger, M., 2008. Managing the risks of climate thresholds:
683 Uncertainties and information needs. *Climatic Change* 91 (1), 5-10.

684 Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data.
685 *Biometrics* 159-74.

686 Lempert R., Sriver R.L., Keller, K., 2012. Characterizing uncertain sea level rise projections to
687 support investment decisions. California Energy Commission, Sacramento, CA, Publication
688 Number CEC-500-2012-056.

689 MacQueen, K.M., McLellan, E., Kay, K., Milstein, B., 1998. Codebook development for team-
690 based qualitative analysis. *Cultural Anthropology Methods* 10 (2), 31-36.

691 Mayer, L.A., Bruine de Bruin, W., Morgan, M.G., 2014. Informed public choices for low-carbon
692 electricity portfolios using a computer decision tool. *Environmental Science & Technology*
693 48 (7), 3640-3648.

694 McDaniels, T., Chang, S., Cole, D., Mikawoz, J., Longstaff, H., 2008. Fostering resilience to
695 extreme events within infrastructure systems: Characterizing decision contexts for mitigation
696 and adaptation. *Global Environmental Change* 18 (2), 310-318.

697 McInerney, D., Keller, K., 2008. Economically optimal risk reduction strategies in the face of
698 uncertain climate thresholds. *Climatic Change* 91 (1-2), 29-41.

699 McInerney, D., Lempert, R., Keller, K., 2012. What are robust strategies in the face of uncertain
700 climate threshold responses? *Climatic change* 112 (3-4), 547-568.

701 Morgan, K., 2005. Development of a preliminary framework for informing the risk analysis and
702 risk management of nanoparticles. *Risk Analysis* 25 (6), 1621-1635.

703 Morgan, M.G., Fischhoff, B., Bostrom, A., 2002. Risk communication: A mental models
704 approach. Cambridge University Press, Cambridge, UK.

705 Nordhaus, W., 2008. A question of balance. Yale University Press, New Haven.

706 Olson, L., Arvai, J., Thorp, L., 2011. Mental models research to inform community outreach for
707 a campus recycling program. *International Journal of Sustainability in Higher Education* 12
708 (4), 322-337.

709 Oppenheimer, M., O'Neill, B.C., Webster, M., Agrawala, S., 2008. The limits of consensus.
710 Science Magazine's State of the Planet 2008-2009: with a Special Section on Energy and
711 Sustainability 123.

712 Robock, A., 2008. 20 reasons why geoengineering may be a bad idea. *Bulletin of the Atomic
713 Scientists* 64 (2), 14-18.

714 Sherwood, S.C., Bony, S., Dufresne, J.L., 2014. Spread in model climate sensitivity traced to
715 atmospheric convective mixing. *Nature* 505 (7481), 37-42.

716 Srivier, R.L., Urban, N.M., Olson, R., Keller, K., 2012. Toward a physically plausible upper
717 bound of sea-level rise projections. *Climatic Change* 115 (3-4), 893-902.

718 Swart, R., Bernstein, L., Ha-Duong, M., Petersen, A., 2009. Agreeing to disagree: Uncertainty
719 management in assessing climate change, impacts and responses by the ipcc. Climatic
720 Change 92 (1-2), 1-29.

721 Tope, D., Chamberlain, L.J., Crowley, M., Hodson, R., 2005. The Benefits of Being There
722 Evidence from the Literature on Work. Journal of Contemporary Ethnography 34 (4), 470-
723 493.

724 Vellinga, P., Katsman, C., Sterl, A., 2009. Exploring high-end climate change scenarios for flood
725 protection of the netherlands. Royal Netherlands Meteorological Institute, De Bilt, WR 2009-
726 05.

727 Wallace, W.A., 1994. Ethics in modeling. Elsevier, Amsterdam.

728 Weitzman, M.L., 2011. Fat-tailed uncertainty in the economics of catastrophic climate change.
729 Review of Environmental Economics and Policy, 5 (2), 275-292.

730 Wood, M., Kovacs, D., Bostrom, A., Bridges, T., Linkov, I., 2012. Flood risk management: US
731 Army corps of engineers and layperson perceptions. Risk Analysis 32 (8), 1349-1368.