Techné: Research in Philosophy and Technology ISSN: 1091-8264
22:2 (2018): 191–229 DOI: 10.5840/techne201862283

Engineering Model Independence: A Strategy to Encourage Independence Among Models

Zachary Pirtle, Jay Odenbaugh, Andrew Hamilton, and Zoe Szajnfarber

Abstract: According to population biologist Richard Levins, every discipline has a "strategy of model building," which involves implicit assumptions about epistemic goals and the types of abstractions and modeling approaches used. We will offer suggestions about how to model complex systems based upon a strategy focusing on independence in modeling. While there are many possible and desirable modeling strategies, we will contrast a model-independence-focused strategy with the more common modeling strategy of adding increasing levels of detail to a model. Levins calls the latter approach a 'brute force' strategy of modeling, which can encounter problems as it attempts to add increasing details and predictive precision. In contrast, a model-independence-focused strategy, which we call a 'pluralistic strategy,' draws off of Levins's use of an assemblage of multiple, simple and—critically—independent models of ecological systems in order to do predictive and explanatory analysis. We use the example of model analysis of levee failure during Hurricane Katrina to show what a pluralistic strategy looks like in engineering. Depending on one's strategy, one can deliberately engineer the set of available models in order to have more independent and complementary models that will be more likely to be accurate. We offer advice on ways of making models independent as well as a set of epistemic goals for model development that different models can emphasize.

Key words: Levins, independence analysis, levees, robustness, epistemic strategy

Zachary Pirtle, Department of Engineering Management and Systems Engineering, George Washington University; 800 22nd St NW, Rm 2670, Washington DC 20052; zpirtle@gwu.edu

Jay Odenbaugh, Philosophy, Lewis and Clark; 0615 S.W. Palatine Hill Road, MSC 45, Portland OR 97219; jay@lclark.edu

Andrew L. Hamilton, College of Natural Sciences and Mathematics, University of Houston; 3507 Cullen Blvd, Room 214, Houston, Texas 77204-5008; alhamil4@central.uh.edu

Zoe Szajnfarber, Department of Engineering Management and Systems Engineering, George Washington University; 800 22nd St NW, Rm 2670, Washington DC 20052; zszajnfa@gwu.edu

> The naïve, brute force approach would be to set up a mathematical model which is a faithful, one-to-one reflection of this complexity. This would require using perhaps 100 simultaneous partial differential equations with time lags; measuring hundreds of parameters, solving the equations to get numerical predictions, and then measuring these predictions against nature. However:
>
> > (a) there are too many parameters to measure; some are still only vaguely defined; many would require a lifetime each for their measurement.
> >
> > (b) The equations are insoluble analytically and exceed the capacity of even good computers.
> >
> > (c) Even if soluble, the result expressed in the form of quotients of sums of products of parameters would have no meaning for us.
>
> Clearly we have to simplify the models in a way that preserves the essential features of the problem. The difference between legitimate and illegitimate simplifications depend not only on the reality to be described but also on the state of the science. (Levins 1966, 421–22)

I. Introduction[1]

Engineers are frequently called upon to develop extremely complex systems and provide advice on techno-social interactions to policy makers. Developing a model is often a required part of building or understanding a system. Scholarship on the role of modeling in policy has shown the importance of reflection on the limits of models, especially when using models to inform contested policy goals in complex, open-ended systems (Sarewitz, Pielke, and Byerly 2000; Oreskes, Shrader-Frechette, and Belitz 1994).[2] The way in which models can inform policy and engineering work depends on the overall strategy of modeling in a field, which we define as a discipline's use of different kinds of models, including implicit modeling assumptions about epistemic goals and the types of abstractions and modeling approaches to be used. We developed our definition of 'strategy' using the late population biologist Richard Levins's 1966 paper "The Strategy of Model Building in Population Biology." The caveats to using modeling in policy imply that engineers should reflect on the overall 'strategy' of modeling that they employ in dealing with complex systems. Some engineers (De Weck, Roos, and Magee 2011) have advocated for multidisciplinary modeling approaches, paying careful attention to social dimensions of technology including policy and regulation variables. While they give some advice on how to model engineering systems, more work is needed to fully assess what engineers' modeling strategies should be.[3]

There can be many possible modeling strategies, but engineers sometimes only see one. For some, the first instinct in trying to solve a complex problem is to create a model of that problem; the more detailed a model, the better. Using phrasing from Levins (1968), we will refer generally to this inclination as a 'brute force' modeling approach, where details are added to the model (akin to adding 'force') as a way to overcome any analytical hurdles. Many students learn from experience or mentors not to go 'too far down the [modeling] rabbit hole,' and sometimes it is suggested, perhaps quoting Einstein, to make a model "as simple as possible, but no simpler." Engineering systems models can differ significantly in focus, from water infrastructure, skyscraper design, supersonic aircraft, to rocketry. The tendency to use more complex and comprehensive models exists in many disciplines: engineering cost analysis (Hamaker 2010); climate modeling (Dessai et al. 2009); and modeling of turbulent fluid dynamics (Argyropoulos and Markatos 2015), among many others.[4] The push to add in more details as a way of responding to problems can influence even an experienced modeler.

We will focus on illustrating one alternative modeling strategy for performing modeling work that can advise modelers engaged with tricky problems. But we must note that there can be many different strategies for modeling. For example, Breiman (2001) focuses on two cultures of data modeling and algorithmic modeling.[5] There can be other modeling strategies that focus on causal structure or historical regression One can also pursue a 'monistic' strategy by using a single, relatively simple model to assess a given problem. Our paper develops one alternative modeling strategy in part to help engineers as they consider the broader set of possible model strategies. By seeing one alternative clearly, a modeler can more fully explore the space of possible modeling strategies. And, while the alternative strategy sketched out below is not the only possible strategy, it is likely to be practically useful in many situations.

Our strategy is an independence-focused, or pluralistic, modeling strategy for engineering systems work, involving the use of multiple, independent models. We show how it may strengthen engineers' analysis to support decision making and design processes by helping to increase confidence in modeling conclusions, and we provide advice on how to implement such a modeling strategy by deliberately creating independent models.[6] Our approach here follows Levins's 1966 'strategy' of using multiple independent models to assess the same system. Our advice focuses on tangible ways to make models more independent from one another.

We develop a framework of types of independence based on a case study of levee analysis that occurred after Hurricane Katrina and explore different conceptual axes for defining independence among models. An engineer can take this list of dimensions of independence, examine the models existing in their domain area, and consider whether there are key areas of uncertainty wherein new, more independent models might be needed. The literature assessing Levins's 'strategy' is applicable to engineering due to similarities that exist in modeling across the sciences as well as to the practical constraints that exist in engineering.[7]

In addition to highlighting the need to consider model strategies more generally, we make both a theoretical and practical contribution by developing a pluralistic strategy in more detail. Theoretically, our approach here also reinterprets and builds upon the scholarship surrounding Levins's 'strategy.' Many scholars have followed Levins and William Wimsatt's approach of assessing the nature of agreement across independent models (Lloyd 2015; Soler et al. 2012; Odenbaugh 2003; Pirtle, Meyer, and Hamilton 2010; Weisberg 2006a; Weisberg 2006b; Vezer 2016; Knutti 2018; Masson and Knutti 2011; Parker 2011, 2018).[8] We contend that the epistemic benefits of using independent models do not require that the multiple models all have the same causal structure, which is a claim made by Lloyd (2015) and Weisberg (2006b; Weisberg 2013).[9] We argue that the core of Levins's strategy is about increased epistemic confidence that arises when *independent* models agree, and show that a study of independence is key to understanding his approach. While there are still ongoing debates about the virtues of using independent models, the general consensus is that there is some increased confidence that results from getting agreement across multiple models; most agree that some type of model pluralism is beneficial.[10] We also make a theoretical contribution by creating a more nuanced typology of how to define independence based upon a case study, demonstrating how independence can be multi-dimensional and part of a spectrum. Practically, the advice we offer on how to define and building types of independence among models can be useful to an engineer struggling to find ways of assessing a given system. We offer a way to operationalize the literature on robustness, putting this conceptual discussion into practice.

In Section II of this paper, we will review some of Levins's original concerns about brute force modeling strategies. Section III is our discussion of Levins's alternative approach of using multiple independent models, which we refer to as a pluralistic strategy. In Section IV, we present a case study of an engineering model ensemble that well represents his independence-focused model strategy. Section V will show dimensions in which groups of models can be significantly independent from one another in terms of their characteristics and intellectual goals. This section offers tangible ways in which models and modelers can be deliberately augmented to try to create an ensemble of independent models. Section VI will conclude with a discussion of advice for engineering modelers.

II. Levins's Criticisms of Brute Force Modeling

The last section briefly discussed the strategy of building a singular, more comprehensive model as a single strategy, which is a commonly used approach.[11] We elaborate on that more here, but we do not claim to do a 'knock down' argument of the approach, which is impossible to do, as modeling strategies may be more or less effective depending on the context but are rarely fully non-viable. Rather, we describe it in order to set up an illustrative contrast with the pluralistic approach to modeling that we'll discuss later in order to help engineers to notice issues they may consider as they choose their own modeling strategies

Based on Levins 1968, we will describe brute force modeling simply as an effort to place all possible relevant variables into a single model, attempting to describe them with maximum resolution and to make predictions over longer timeframes. The aim is to find a model which maximizes generality, realism, and precision. Recent examples of brute force modeling include efforts to increase climate model accuracy at local levels on 100-plus year timeframes, agent-based models of large populations, and turbulence models of supersonic airplanes (Dessai et al. 2009; Shackley et al. 1998). Levins's (1966; 1968) original claim about brute force modeling was made in response to the biologist Kenneth Watt, who was advocating a strategy of creating one massive model that could attempt to model all variables simultaneously, including both ecological variables (such as predator/prey totals) and genetic information. These variables all interrelate but can change on fundamentally different time-scales.

The boundary between brute force modeling and 'normal' modeling of a system—wherein one adds and inserts relevant details but still retains many key idealizing assumptions and simplifications—can be unclear. One alternative to a brute force model would be to create just one single model of low to moderate comprehensiveness; one might label this approach a 'monist' approach, as it involves a singular but not unduly burdensome model approach. The general difference between a brute force model and a relatively simple 'monist' model can change as time goes on and the overall state of the art and capability of modeling approaches and scientific understanding matures. Nevertheless, the examples of brute force modeling cited above help show that this approach does commonly occur. Brute force modeling, like monist approaches to modeling, can also be successful in certain contexts, particularly if the bounds of a system are well-defined and there is a strong ability to iteratively test the model.[12]

Levins gives three reasons why brute force approaches to ecology would not work in 1960s ecology and generally; his reasoning is applicable to models of many different kinds of complex systems.[13] Levins's three main arguments are:

> (a) There are too many parameters to measure; some are still only vaguely defined; many would require a lifetime each for their measurement.
>
> (b) The equations are insoluble analytically and exceed the capacity of even good computers.
>
> (c) Even if soluble, the result expressed in the form of quotients of sums of products of parameters would have no meaning for us. (Levins 1966, 422)

The lasting significance of Levins's argument against a brute force approach has been debated in the literature (Odenbaugh 2003; Odenbaugh 2006; Weisberg 2006a). The problem of data measurement, "(a)," for a wide range of topics is a still difficult and acknowledged problem in many fields (Sarewitz, Pielke, and Byerly 2000). Data availability will likely always place some limit on how much data can go into a brute force model. The greater the number of distinct parameters, the more data one needs to estimate values of those parameters. Levins's claim in "(b)," that good computers cannot solve many of the problems of population ecology, is still relevant in a way. Computing capabilities have increased to the point where numerical methods techniques can make approximations that solve complicated, analytically insoluble equations. As has been alluded to, the boundary between what is 'brute force' modeling versus more tractable modeling is historically contingent and evolving and is affected by computation (Levins 1993). Regardless, the amount of realism underlying new, extremely complex models and their predictions can be difficult to determine.

Levins's criticism "(c)," that model results may "have no meaning for us," questions the comprehensibility of a model, and highlights a practical risk that is exacerbated when following a brute force approach. One way in which a model prediction can lack meaning is if it is impossible to provide a causal explanation for why the result occurred: lacking such a story, there is still a very important sense in which we do not understand the output result. Paul Humphreys (2009) has labeled this challenge the 'epistemic opacity' of models, where humans are not able to follow the logic of the model. While computer models have increasing power and can produce predictions for very complex systems, the difficulty of explaining why a result occurs persists due to the difficulty of explanation using complex simulations (Winsberg 2010). Simulations exacerbate this problem since it can be challenging to make sense of programming code especially if one did not build the model themselves. Weisberg (2006a) argues that the crux of Levins's argument in "(c)" may instead be that within a complex model it can be impossible to identify exactly what the causal process is that leads to a given result.[14] All models, even simpler models used in the alternative strategy described below, present these issues, but the challenges increase as a model increases in complexity. A brute force strategy may by its very nature complicate the ability of a modeler or stakeholder in a model analysis to clearly assess model causality due to challenges in epistemic opacity.

III. Pluralistic Modeling as Strategy: "Truth Is the Intersection of Independent Lies"

Out of many possible modeling strategies, one can embrace multiple ways of assessing a system as opposed to focusing on a monist or brute force modeling approach. Levins specifically embraces the idea of using multiple independent models, which can be a helpful way to assess complex systems. We call this approach a pluralistic strategy because it acknowledges that multiple partial approaches can yield value. Part of the motivation for a pluralistic approach follows from the faults identified in the brute force modeling strategy, which reflects a tendency to make a model increasingly more comprehensive. Adding more details to a model raises challenges in having the right data and in comprehending the model, which creates a tension between the comprehensiveness and the overall realism and accuracy of the model. Levins has a provocative way of describing the tradeoff between model details and model realism. He writes: "No single model can meet all the requirements of generality, realism, precision and manageability. Therefore we need a cluster of models" (1966, 304). Subsequent authors have acknowledged and emphasized this tradeoff (Odenbaugh 2003; Weisberg 2006a). We will return to these model requirements later, as they help lay out dimensions in which models can be deliberately engineered to be independent from one another.

Levins then sets out to describe three independent models that he uses for doing analysis.[15] Levins lays out the virtues of using a multi-model ensemble as follows:

> [W]e attempt to treat the same problem with several alternative models each with different simplifications but with a common biological assumption. Then, if these models, despite their different assumptions, lead to similar results we have what we can call a robust theorem which is relatively free of the details of the model. Hence our truth is the intersection of independent lies. (Levins 1966, 423)

The virtue of using independent models is straightforward. Utilizing independent models allows one to test conclusions from those models against one another, which can provide a source of confirmation when the models agree with one another. Levins then goes on to individually discuss his ecological models that make assumptions of differing levels of realism and vary in their level of quantification and precision, but all of which are based in evolutionary theory. The models are different from one another because they have distinct characteristics and assumptions, with two focusing on ecological assumptions and only one including genetic assumptions. Levins claims that the ability of the models to agree upon results can lead to a "robust theorem," where a shared claim is more likely true. A group of models could lead to a more-likely-true prediction for a variety of model outputs, including numerical or qualitative claims, as well as more accurate explanations of why things happen in the modeled systems. This concept of multiple-determination of results has also been referred to as 'consilience,' though the literature treats the concepts of independence and consilience in different ways.[16] The search for independence among groups, or ensembles, of models can thus aid in making shared model results more accurate, with the results potentially leading to a number of scientific goals such as better predictions, explanations, representations, or having tools for intervention (Waters 2004). Levins listed out criteria for what models should focus on, including how realistic, precise, and general the models are. Section V of paper will expand upon? and operationalize the dimensions of the model-building strategy that Levins developed.

Levins did not develop a panacea: the multiple, independent model strategy that Levins advocates is helpful but some deeply complex systems may conceivably defy any. Indeed, one may apply a multiple model strategy to a given problem and get no meaningful agreed-upon results, despite the costs incurred from using multiple independent approaches. Moreover, all models can be subject to some of the challenges discussed above, such as epistemic opacity, ability to get data to handle parameters across multiple timelines, etc. An independence-focused strategy can be fallible and should not be seen as automatically better than a brute force modeling approach. As mentioned earlier, brute force modeling approaches can work well in some systems, such as aircraft design, where experimentation and iteration of design are possible. There is a critical difference between how epistemic opacity problems apply to the two approaches, as many of the challenges of modeling are uniquely exacerbated when one pursues a more brute force approach. In contrast, using multiple models can reduce the inclination to double down on the details of any one model, thus making some of the problems associated with the brute force model approach less severe.

Again, one may want to pursue a myriad of modeling strategies, but it is worth reflecting on the virtues of a pluralistic modeling strategy versus a brute force or simple monistic strategy. The limits of when brute force modeling approaches can work is often uncertain, and pursuit of a brute force modeling approach can be counterproductive when dealing with policy and other complex decisions, including those involving value disputes. In situations where it is important for humans to have a causal explanation of how the brute force model works, it is likelier that a pluralistic or a monistic strategy may be better to pursue. However, there is still a need to see what modeling strategies work better empirically and to question whether one strategy might have more epistemic limits inherently.[17] A pluralistic strategy is likely to be preferred over a simple monistic strategy when there is sufficient uncertainty such that one believes multiple disciplinary lenses to assess the system are needed, which may imply multiple or more comprehensive models are needed.

To show how Levins's idea of model independence can be a more general modeling strategy in engineering, in the next section we will reflect on an engineering analysis in which using independent models was key to getting confidence in a conclusion.

IV. A Pluralistic Case Study: Three Models Used to Assess a Levee Failure

To develop our framework, we discuss a case where three different types of engineering models were used. In the wake of Hurricane Katrina and the numerous levee failures in New Orleans, the Interagency Performance Evaluation Taskforce (IPET) systematically analyzed the numerous levee and floodwall failures using a set of models (Interagency Performance Evaluation Taskforce 2007). Because each model might be fallible on its own, the IPET used multiple models, which agreed upon a cause for failure. These included: a finite element analysis model, a limited equilibrium model, and a physical centrifuge model, each of which assessed failure at the 17th St. Canal, and each independently agreed on the same cause of failure. The 17th St. levee wall failure was particularly important because it fell prior to being overtopped, and it raised questions about whether there had been faulty designs or whether there were failure modes for other levees that had not been addressed.

We will review these models to assess whether the strategy for model building (either deliberate or implicit) in this levee failure analysis is akin to a 'brute force' or a pluralistic modeling strategy.[18] Understanding the forms and degrees of independence between the models is a conceptual challenge that helps define a field's overall modeling strategy. The extent to which a field's modeling strategy is pluralistic depends on how different the multiple models are from one another across several dimensions that we will discuss in the next section, inductively drawing off of the details of this case study. Parts of this section were initially developed in Pirtle 2010 and have been significantly expanded here.[19]

In general, the IPET report's conclusions are generally seen as technically sound, though some have challenged the analysis and some of the IPET team's composition and data assumptions. Overall, it is clear that the shared model agreement among multiple models was key for the IPET team's belief that they were correct in their determination of the cause of the 17th St. levee wall collapse. The IPET team's models all supported a conclusion that failure occurred due to a gap emerging between the levee wall and the water that allowed pressure to be applied against a broader portion of the wall which, combined with an area of weak clay soil at the base of the wall, caused the levee to fail. This explanation of a gap exacerbating failure was seen by the IPET team as a novel and unexpected contributor to failure, with some commenting that this type of failure was not seen in the textbooks and that this failure mode needed to be investigated in other systems (Kestenbaum 2006).

There were some common assumptions that went into all of the levee models as they were based on the same initial datasets. The IPET engineers had post-levee failure information collected from the site along with rough knowledge of floodwater height at the time of failure. The team had a budget greater than $20 million and involved over 150 experts, some of whom were from independent universities but many of whom worked for the US Army Corps of Engineers (USACE), which had led construction of the levee walls decades earlier. There was collaboration across the IPET sub-teams along with the shared data assumptions, both of which were later criticized. The models discussed here were used to evaluate all levee failures in the aftermath of Hurricane Katrina, but they were critically important in evaluating the 17th St. Canal levee failure (Seed et al. 2006). Each model approach eventually led to a shared conclusion about the likely cause and location of failure for the 17th St. levee. The reports' conclusions were peer reviewed by groups from the American Society of Civil Engineers and the National Research Council. Some studies did disagree with and criticized the IPET, which will be discussed later.

The following models were used:

Limit Equilibrium Assessment Model (IPET 2007, V-41-V43): Limit equilibrium analysis (LEA) is a relatively old method of analysis of the stability of slopes and embankments developed by civil engineers, valued for its simplicity, accuracy, and ease of computation (Duncan and Wright 2005). LEA analyzes possible failure along a slope by positing a failure plane. Given the known loads caused by the weight of the water and of the soil, moment and force equilibrium are applied to establish the forces along the postulated shear plane. A shear strength is assigned to each portion of the postulated failure surface based upon assumed strength parameters and the applied normal forces. A factor of safety is calculated as either the resisting moment about the center of rotation divided by the driving moment or, for sliding block failures, the integrated resisting shear strength divided by the integrated applied shear stress. In a LEA analysis, all possible ("kinematically admissible") failure surfaces must be evaluated to find the surface with the lowest factor of safety. By definition, a factor of safety less than one on any surface means that the applied shear stress along that surface is great enough to exceed the shear strength along that surface and cause slippage across that LEA failure plane.

Analysis of failure at the 17th St. Canal was done by examining what conditions would generate a factor of safety less than one. The LEA model indicated that for the factor of safety to decrease below 1, or in order for the model to predict failure at the flood-level heights seen during Katrina, it was necessary to assume that a gap emerged between the floodwall and the canal side soil. The inferred cause was that the gap exposed the water to more of the wall, allowing pressure to be applied at lower and lower levels, where eventually an area of weak soil (clay) at the base of the wall started to give way laterally at around 6.5 feet of flooding.

Finite Element Analysis Models (IPET 2007, V-45-V-52): Finite element analysis is a commonly used practice in engineering that subdivides an object into discrete 'elements,' which can be one-, two- or three- dimensional beams or blocks. Finite element analysis is in some ways merely a method of solving differential equations of force equilibrium, conservation of mass, and continuity of displacement collectively across the domain being modeled for the specified boundary conditions. The accuracy of a FEA is highly dependent upon the grid resolution employed to model an observed system results depend upon both the boundary conditions and the inputs to (demands upon) the system. In the IPET report, FEA models were used to generate factor-of-safety values, which are the calculated maximum yield strengths of the soil divided by the stresses calculated in the FEA model for the flood conditions along an assumed failure surface. Over 3000 element cells were used across two dimensions in order to assess deformation, with the strength of the elements being determined by the same soil strength data used to inform the other models. The factors of safety in FEA analyses conducted by the IPET team were all above one, indicating that failure would not occur, except in the case of models which assumed the creation of a gap in between the flood wall and the canal embankment. Because the gap exposed an area of weaker clay soil, the model assessed how weak soil strength at the base of the wall could combine with a gap between the wall and the water. Both the weak layer and the gap were needed for the FEA model to predict failure prior to being overtopped. Thus, there was some agreement between the FEA and LEA models.

Centrifuge Models (IPET 2007, V-43-V-45): Physical centrifuge models attempt to replicate at a small scale the performance of large geotechnical systems, e.g., flooding of the levee. Because of difficulties in physical modeling of the in situ geometry and exact properties of the ground, the strength of centrifuge testing lies primarily in identification of mechanisms of failure, as well as in calibration and validation of numerical models, i.e., by numerically modeling the centrifuge test. A centrifuge model test of an idealized levee system was conducted by the IPET team (IPET V-43). Materials with strengths close to expected levee strengths were used, with a preference for choosing materials that were well understood in centrifuge testing, such as Nevada sand.[20] In a similar way to the FEA and LEA models, the centrifuge model showed the formation of a gap between the water and the wall that formed, which then led to slippage on a lower level of clay. This thus agreed with the other two models, but it predicted the gap formation in a stronger way (i.e., having it happen) than merely assuming that a gap foundation was necessary in order to get the model to predict a failure. The model likewise suggested that the location of failure was along a layer of clay at the bottom of the structure, meaning that all three models were in agreement.

A. How Using Independent Models Increases Confidence

To synthesize the results cited above: the 17th St. Canal breach was analyzed by IPET using three different models. Each of the models used initial conditions (levee cross section size, soil properties, etc.) based upon observed data from the field. Thus, in a certain sense, the models were based upon the same initial data assumptions, and in that respect were not significantly independent from one another. However, the underlying analytical principles behind each model are significantly different from one another. The principles for the models at hand are:

- FEA models solve equations of stress and strain (which incorporate more physical phenomena than the LEA static equilibrium analysis) across small elements throughout the levee.
- LEA solves static equilibrium equations over an aggregated plane where failure is assumed to occur, determining a factor of safety for a given load.
- Centrifuge models are idealized physical models subjected to similar loading as experienced in the field, with the experiment results used to identify the governing mechanisms of behavior in the field.

The physical model is perhaps fundamentally different from the mathematical models, and there is some literature that explores the different types of knowledge that physical 'things' can carry in contrast to numerical models (Baird 2004). While Levins's original argument for using multiple models was only applied to mathematical models, the concept is easily extended into thinking about physical models.[21] The LEA model solves broad equations of static equilibrium along assumed failure slopes, which thus avoids calculations of strains (and only calculates stress on the postulated failure surface) and does not attempt to understand the integrated system behavior throughout the levee. The FEA model includes static equilibrium equations but is more comprehensive in its stress analysis; however, given its greater resolution and more detailed input soil property requirements, it is perhaps more subject to errors in soil strength calculations.

The narrative of the IPET report strongly highlighted the agreement of each model on there being a gap between the wall and the levee and that a weak soil layer was needed for failure to occur, which they then confidently decided was the explanation for the failure. They advocated a deeper study of gap formations as a contribution to future failures. The IPET team's implicit argument is that the agreement of the three models upon a shared conclusion gave reason to believe that the shared conclusion was more likely true. This is an embodiment of Levins's conceptual framework, which provides several ways to characterize the situation. How independent are the models from one another? They each use different levels of resolution, with the LEA model likely being the most abstract, as fewer details of the system are modeled. The FEA model was relatively less abstract in terms of its modeling many different individual elements, and the centrifuge model had a scaled down physical replica of the levee wall, which reflects the system in a different way. The analytical weaknesses of each model do appear to be different based upon how the models work. No set of models can be completely independent (or different from) one another (Goodman 1970), as they all involve some common understandings from physics and other characteristics. The differences between these models was enough to warrant increased epistemic confidence as a result of a robust agreement between what are largely independent models. When independent models, each with their own inadequacies and idealizations, agree, their agreement significantly increases one's confidence in model results.

There are strong criticisms of the IPET report which should be considered, which complicate its use as an exemplar of independent analysis. Some believe that the IPET used incorrect data: Berkeley professor Bob Bea argues that there was a weak layer of clay soil located higher along the levee wall which was responsible for failure and that a simple, negligent design flaw likely caused by the USACE was responsible (Kestenbaum 2006). This claim can be read as implying that the models actually were not sufficiently independent from one another, as they relied on the same initial assumptions. The IPET report addresses this claim, disagreeing with Bea's evidence for a weakness that should have been detected, and recent publications have backed up the IPET's overall conclusion on failure mechanism.[22] Others were concerned that that the IPET team had too strong a membership from the USACE, which was responsible for the original levee design. The criticisms here are akin to arguing that there was not sufficient independence among those performing the analysis and reviewing it. In both of these criticisms, one way of revising the IPET's work would have been to push to create more independence in data assumptions and team membership.

This levee failure analysis is a good engineering example of what a pluralistic modeling strategy can be, which was especially effective here given a complex problem with incomplete data.[23] An alternative brute force modeling approach could have ignored the push for multiple models, although that likely does not make sense as a conceptual strategy.[24]

V. Dimensions of a Strategy; or, Knobs to Vary to Create Independence Among Models

This section clarifies several of the ways in which one can deliberately engineer one's models to be independent, recognizing that there are situations where no modeling strategy can work.[25] By inductively reflecting on the above case study, we will discuss the two main areas for describing one's model-building strategy: first, diagnosing what independent principles and methods underlie the models; and second, assessing the epistemic goals of a model. We define our notion of model independence as non-binary and existing on a spectrum where models can be more or less independent, but with some cases such as the levees example embodying a significant amount of independence. Complete independence is not possible because, as Giere and Goodman note, everything can be seen as similar to something else in some respects and to some degrees.[26] The following is a set of key dimensions (or 'respects') in which models can be independent (or dissimilar) from one another. A core part of our contribution is developing the below framework of independence. Other literature uses the notion of independence both directly[27] and indirectly,[28] but not synthesized in a way to provide advice to modelers. The value of independence along each of the above dimensions of independence needs to be carefully assessed and thought through. The pursuit of independence among models is an art—its cost must be weighed by the model builder/user to assess what areas of uncertainty need most to be addressed. A group of experts with sufficient accountability to the public and an understanding of the ways models may be independent from one another is likely best positioned to determine how much independence is needed.

A. Identifying Independence among the Models/Modelers: Physical/Social Characteristics

Differences in the physical and social characteristics of the models and modeler groups are the first set of definitions of independence that we explore. Differences among the modelers themselves also can be key, as the modelers shape a model using their own experience and tacit knowledge. Drawing on Pirtle 2009 and Pirtle, Meyer, and Hamilton 2010, we have elaborated on a set of dimensions by which models/modeler groups can be independent from one another. Four general categories are used:[29] causal mechanisms underlying representations, who is involved, what is represented, and the knowledge context contributing to the model creation and use. We will describe these and then later show how independent the levees models are in these regards.

'How': *Causal mechanisms and physical principles underlying representation.* The underlying assumptions about how the system works can differ from model to model. Many models will agree on fundamental issues such as principles of conservation of mass and energy (both of these underlay the LEA/FEA models described above). A physical model like the centrifuge, being embodied in the world, may draw upon unknown features that a model designer may not be aware of. Hacking (1985) refers to the epistemic power of using multiple microscopes to observe the same phenomenon when the instruments are based on "unrelated chunks of physics"; the odds of seeing a false phenomenon through physically different perspectives is much lower.

'Who': *Historical and sociological perspectives*. Modelers' training and working approaches can encourage shared strategies, which can actually decrease independence across models. The epistemic culture behind a modeler's approach and the goals of their research can be critical (Knorr Cetina 2009). At the same time, collaboration can increase the quantity of critique of ideas and the proliferation of different approaches. Having models created based upon another program's source code and approach is a source of interdependence. Shared history among modelers can affect all other parameters, as debates about the meaning and nature of different assumptions can vary significantly. It is likely best to think about independence among models and modelers jointly, to capture the broad interaction between the analyst and the model. In the IPET team, the models themselves shared some common history as being used in geotechnical engineering and the team members themselves had broad social connections to the USACE and other organizations that had long used and developed those tools.

'What': Aspects of Representation. *Parameters*. Specific assumptions about a physical system that are below the level of a general scientific principle can often be distilled into parameters. These are a key part of how physical principles are incorporated into a model. For example, assessing the yield strength of soil will result in a parameter that can be input into the FEA/LEA models described above. Parameters are specific variables which can be input into code that reflects more general scientific laws (which define the equations used in the code). The centrifuge model used materials with strengths close to but not exactly similar to the strengths assumed in the models.

Scope and idealizations: The way that models resolve the system can be different. This can involve models that use different system boundaries and resolutions, or that make different idealizations which adjust features of the real-world system for ease of modeling. FEA models can break down the system into a set of discrete chunks, while LEA models only seek to assess a single failure plane. Having different scope and idealizations can also include models that operate at different levels of abstraction, wherein one model may have many details of a component of the system and another model may treat it at a significantly abstracted and simplified level (Abbott 2007).[30] It may be the case that having different models at different levels of resolution is important to pursuing a Levins-like epistemic confidence for an agreed upon claim.

Numerical coding schemes: If physical equations are resolved through different coding schemes or software packages, this can be a source of independence. Truncation errors can aggregate differently over space and time based on the numerical coding approach chosen. Given that the models used in the levee case were so different, this was not an issue. Further, other software packages were used to replicate the FEA and LEA analysis models, which confirmed the results there.

Knowledge Context for Model Creation: *Data*. Using different data, in terms of variations in physical sources of data as well as locations and time periods used, can be one way for model analyses to be independent from one another. The use of different data does not occur in the IPET models, as all were developed after the levee failure and using the same post-failure data set. The Bea criticism is that there was an additional area of soil where a problem occurred, and that this data was missing from all three of the IPET models. Edwards (2010) explored the role that models of data can play in informing larger, more comprehensive models, highlighting the complexity of assumptions used in many kinds of models.

Again, if one wants to pursue a pluralistic modeling approach, then one can choose to deliberately build independent models that have different assumptions and characteristics along the above dimensions. The levee models were most independent from one another in their use of different scope/idealization of the system, though the physical model may draw on separate physical principles as well. If one wants to explore having additional independence among 'who' is involved and their associated history and disciplinary approaches, one can draw in new analysts with different histories, backgrounds and institutional affiliations to perform or verify analysis. Using different timeframes and types of data can help show that one's theoretical understanding in the model and associated system is accurate.

B. Epistemic Desiderata of Models: Building Models that Focus on Realism, Precision, Generality, or Manageability

We will now introduce new dimensions of independence that tie to the goals of a model. The dimensions of independence noted above are physical and social characteristics of models and the modelers who use them, which collectively are relatively tangible and definable. These dimensions do not deal with epistemic qualities of a model, where 'epistemic' refers to the way in which the model helps to create knowledge and what the nature of that knowledge is. Levins called these epistemic qualities of a model "desiderata" and discussed them independently from the ‘assumptions’ within a model. The four desiderata that he explicitly mentions are generality, realism, precision, and manageability: Do we believe the models are very realistic in nature? Do they have as much precision as is possible? How generalizable is the model to other systems? Is the model manageable to think through and to practically operate? All of these are epistemic qualities in that they are tied to the overall knowledge embodied in the model and to what kinds of knowledge and claims that the models can be used to create. In contrast to these epistemic desiderata, the earlier dimensions of independence are more akin to characteristics that the models/modelers bring to the table (almost akin to being ‘inputs,’ with the desiderata being about the process to create outputs). The epistemic qualities of a model can only be assessed holistically, considering all aspects of a model and the state of the existing science.

There are practical reasons to care about which model desiderata one focuses on, as Levins argues that these model goals can be in conflict:

> It is of course desirable to work with manageable models which maximize generality, realism, and precision toward the overlapping but identical goals of understanding, predicting and modifying nature. (Levins 1966, 422)

Levins argues that trade-offs must be made when choosing between different goals, which he calls desiderata, and that one model cannot achieve every goal.[31] Having multiple models that focus on different desiderata can reduce this conflict. The main reason Levins states that there will be a tradeoff in focusing on specific goals is due to limits in what any one person or group can feasibly do in a given moment of time. Levins does not prescribe limits to how many desiderata or epistemic goals one can have for a model, and views his four desiderata (generality, realism, precision and manageability) as non-comprehensive, meaning that additional desiderata should be addressed.

We will elaborate on each model desiderata here as additional dimensions by which models could be made independent from one another. A pluralistic modeling strategy can have different models/modelers focus on different goals for building their model, with some focusing on encouraging realism and others focusing on manageability or precision. When surveying an existing set of models, one can assess how independent the models are from one another in these epistemic goals. Our definitions here are inspired by Levins, but we build on his definitions significantly.

Realism: There are three different ways to conceive of realism in a model: as the inclusion of all the relevant causal factors; as the exclusion of false assumptions; and as accuracy in depicting the system, including in making predictions. Including more variables and excluding false assumptions both increase the realism of a model. These two objectives are related but separate: suppose you are modeling a block sliding down a plank (a classic mechanics problem). You can include the force of gravity. Sometimes you make simplifying assumptions that the plank is frictionless. Others exclude the friction of the air against the block. As Giere (1999) notes, there is no limit to how many causal variables one could ostensibly include—why not include the infinitesimal effect of the moon's gravity on the block? Including more variables increases realism, to a point.

The second aspect of realism, avoiding false assumptions, can be seen in how fluid mechanics models flow through a pipe. These models almost always assume that the liquid itself is a continuum. This is strictly a false assumption: the liquid is composed of atoms, and of the invisible spaces between the atoms.[32] Sometimes, including a deliberately false assumption is the only way to make a problem tractable. In many cases this assumption has no effect on the accuracy of a model's predictions but is nevertheless not true.

The last dimension of realism is accuracy in predictions, which can only be assessed through experimentation and testing. This dimension is often what managers care the most about when reviewing the results of a model, but for some long-term or sufficiently complex predictions it can be difficult to assess how accurate a model is (Sarewitz, Pielke, and Byerly 2000).

Precision: this dimension deals with different ways in which a model can provide detail about what is happening in the world. We will describe four aspects of how it can manifest in model design/model operation. The first definition of precision is the ability of the model to provide the same answer given the same initial conditions, which is a classic definition of precision. A second definition relates to the resolution of a model. Low-resolution models may not divide a system up into as many parts or element, whereas a high-resolution model can have many thousands of elements. Predictions for high-resolution models can provide predicted 'micro'-details about behavior in a system. In finite element modeling, newer models tend toward higher resolution in an attempt to identify more precise and narrow effects within a system. In such models, one can assess accuracy of the results at both the macro-level (i.e., did the levee break?), and also at the micro-level (i.e., did the levee break at the exact same point that the model predicted?).

A third element of precision is the ability to describe uncertainty. Some models do not include uncertainty estimates and merely produce point-estimate predictions of system behavior, whereas other models produce uncertainty intervals underlying each predicted variable. Models that incorporate random chance within their operation can describe uncertainty in a different way: a Monte Carlo simulation can be used to run the model numerous times and by the frequency of observed results predict the probability of a given result occurring in reality.

Lastly, a fourth aspect of precision is the degree of quantitativeness of a model, which is a vague but important idea. When a geotechnical engineer does a limited equilibrium analysis of a possible levee failure, she will calculate the stress on the levee and see if it is greater than the yield strength of the soil along an assumed failure slope. If actual stress is greater than the yield strength, or the Factor of Safety is less than one, then the engineer will know that deformation and failure will occur. If the engineer goes further to quantify the predicted amount of deformation (which the LEA model cannot do, but the FEA model can), this is a more 'quantitative' conclusion than the simple qualitative estimate of whether deformation or possible failure will occur. However, quantitative models are not always more desirable than more qualitative models (Puccia and Levins 1985).

Manageability of a model is another key consideration. The speed with which a model run can be performed on a given platform is a reflection of the computational intensiveness of the program. Faster run speeds can make the model easier to use. Another key dimension of manageability is the comprehensibility of a model and how well a human can understand the model. This can involve several traits, including ease of use and comprehensibility of results and understanding of the reasons why the model behaves in a given way. The epistemic opaqueness of a model, as described above (Humphreys 2009), is thus tied to the manageability of the model. Levins suggests that brute force models can be too complex for a human to understand the meaning and significance of results, which we would characterize here as a manageability concern.

C. Framework for Encouraging More Independence using the Levees Case

To make the above forms of independence seem more tangible, we review how they apply to the levee case and describe how one could try to deliberately change the models to have more independence. Models and modelers can be independent of one another by being different in their physical/social characteristics as well as in the epistemic goals they focus on. Table One below lists the different dimensions of independence noted in the prior sections, then applies these dimensions to the levee case and discusses the extent of the models' independence from one another. The question of how to build (or 'engineer') more independence into the set of models is addressed in the last column. In some dimensions, the models listed below already represent fairly independent forms of analysis.

In some cases, as noted in the table, it is unclear how more independence among the models could be achieved. In assessing the degree of independence among models, there are likely not to be simple, easy-to-assess 'yes or no' binaries that of whether models are independent from one another. The dimensions here lay out the different types of independence that can matter for a modeling group. One can create a similar table for other cases or groups of models with the dimensions of independence applied to the models under examination. Such an analysis can help to determine how the models differ and what additional type of independence might be useful. A modeler can develop a qualitative sense that one group of models may have significantly more independence than another group, or see areas where a group of models lack meaningful independence.[33] The levee

Table One: Dimensions of independence and knobs to increase independence, applied to the case study.

Category of Independence	**Specific Dimension of Difference**	**By Dimension, How Independent Are the Levees Models from One Another?**	**Ways to Engineer More Independence among the Models**
Characteristics of the Modelers/Model Teams			
How: Causal Mechanisms Underlying Representation	Basic Physics. Core causal mechanisms that underly the model.	Uses significantly different ways of resolving the system across FEA and LEA, with physical model involving its own embodied assumptions and behaviors. Key principles include: LEA: Calculates slippage across a failure slope. FEA: divides the system into many different elements, with stress and deformation occuring in each one.	Already seems largely independent, unclear how to become more independent than having physical and numerical models alongside different ways of resolving the relevant phenomena.
Who:	Historical/Sociological Perspectives	Historical separation of the tools is hard to establish as they're from a shared disciplinary past. Sociologically, most IPET team members worked for US Army Corps of Engineers with other independent members, alongside outside peer reviewers.	For historical differences in the tools themselves, have a different disciplinary group model and assess the same situation using analysis approaches outside of geotechnical engineering. Sociologically, having analysts be less connected to the US Army Corps of Engineers might lead to more independence.

Category of Independence	**Specific Dimension of Difference**	**By Dimension, How Independent Are the Levees Models from One Another?**	**Ways to Engineer More Independence among the Models**
What: Aspects of Representation	Parameters	Each model parameterizes slightly different phenomena, but all of it is tied to assumptions of soil strength and assumed height of water at over-topping. FEA assigns yield strength to each element. Arguably there are no parameters in the centrifuge model although materials are chosen to have strengths close to expected strengths of the actual soils; scaling parameters are used to calculate size and centripetal force used.	Already seems largely independent.
	Scope and idealizations	This may be the area of greatest independence among the models, with each resolving different parts of the system. LEA abstracts the system into a series of failure slopes, FEA idealizes it into a series of homogenous grids, and the physical model may idealize in a very different way.	Already seems largely independent. Different grid sizes could be used, as well as simplified abstracts of each type of model.
	Numerical coding	Not applicable here as the models are so different that they do not attempt to code the same assumptions/ parameters. Alternative LEA/FEA software packages were used to spotcheck the analysis.	Not as applicable in this context.
Knowledge Context	Data	Little independence as each model used the same data to inform its analysis.	One could use different datasets to inform the analysis, perhaps assessing the data at different levels of abstraction.

Category of Independence	Specific Dimension of Difference	By Dimension, How Independent Are the Levees Models from One Another?	Ways to Engineer More Independence among the Models
Epistemic Goals:			
Realism	Comprehensiveness: includes as many features as possible	Models differ in how comprehensive they try to be. The LEA definitely tries to be less comprehensive than the FEA model. The centrifuge model, being physical, may be said to be much more comprehensive in terms of capturing actual features of the soil-structure interactions.	Similar to the scope category above, models already seem independent. One could add more grid cells into the FEA or make the centrifuge model even large and incorporate more nuances into that model.
	Idealizations (avoids false assumptions)	It is unclear how the models differ in their attempts to minimize idealizations. The LEA model idealizes the system by ignoring many features but the FEA model introduces new assumptions/idealizations by creating new grid cells. (Ergo, each grid cell is assumed to be homogenous). Beyond scaling, it is unclear what idealizations are made in the centrifiuge.	Already seems largely independent.
	Accuracy—result is believed to be true	All three models supported a common claim about cause of failure, although their individual predictions varied in whether they indicated failure would occur under assumed conditions vs predicting exactly how failure occured. Future work would criticize assumptions on strength but most agree predicted failure mode likely is accurate.	Accuracy/confidence in new, hypothetically independent models is hard to assess a priori.

Category of Independence	Specific Dimension of Difference	By Dimension, How Independent Are the Levees Models from One Another?	Ways to Engineer More Independence among the Models
Precision	Repeatability of result; granularity of results; discusses uncertianty; quantitativeness of results.	The models differ in ways in which they can provide more resolution and information about uncertainty. The LEA model shows where the failure slope is likely to be, but cannot describe the motion of the system as FEA and centrifuge can. None do a good job of capturing uncertainty bounds in the final analysis.	Already seems largely independent, but could do more to capture uncertainty.
Manageability	Run speed; comprehensibility of the result.	The models differ in the amount of effort needed to set up and run the model and assess its input. LEA is simpler and easier to manage than the other two. The centrifuge actively predicts the formation of the gap whereas the other wo models can only explain failure if they assume a gap occurs, which may make it more understandable.	Already seems largely independent.

example had significantly independent types of models in terms of the causal mechanisms underlying the models and aspects of representation, but the overall IPET effort was criticized for a lack of independence in other areas, including the lack of social independence in terms of having different teams perform the modeling analysis as well as the use of similar data assumptions on soil strength in each model. Regardless, criticisms of the 17th St. levee failure analysis hinge on different aspects of independence as discussed here, including a potential need for more independence.

VI. Engineering Model Independence

Getting clear about what models, and model ensembles, can and cannot do should be an important part of the agenda for better using models in policy-making and the creation of complex systems (Sarewitz, Pielke, and Byerly 2000). In Levins's phrasing, every field implicitly has its own strategy of model-building which affects how models are used in design and policy. Given the pressures and time constraints of professional life, it is possible for a discipline's model-building strategy to be more ad hoc instead of deliberately structured.[34] We laid out one possible modeling strategy that focuses on using models that are meaningfully independent from one another. Independence among models is a complex issue involving multiple social, technical and epistemic dimensions. We laid out a taxonomy of types of independence, drawing from a case study of the models of the levees at the 17th St. Canal as a case study for thinking through what an alternative modeling strategy might look like. The criticisms of the levee analysis can be seen as a call for additional independence in terms of social actors and the types of data used. On the whole, reliance on multiple, independent models served to help the IPET team become more confident in their analysis.

A modeler can explore new strategies of their own, and sometimes employing a brute force modeling strategy is desirable. As noted above, issues such as computing power that affect the desirability of a brute force modeling strategy can change over time, and likely is tied to our ability to comprehend the complexity of a system. When a modeler cannot give a clear causal story using a brute force model (Winsberg 2012; Humphreys 2009), it may be beneficial to use a pluralistic strategy. When using modeling to inform policy decisions that have value disputes embedded within them, the pluralistic strategy's multiple, simpler models may help make the analysis more transparent to decision makers, improving the salience of the analysis (Sarewitz, Pielke, and Byerly 2000).

If a modeler or group decides it is worthwhile to implement a pluralistic modeling strategy, our advice is to engineer independence into the building and use of one's models. This advice is based on the assumption that the more independence there is among models, the greater the confidence in the overall results. If one wants to deliberately engineer independence into a group of models, then varying the assumptions of models as well as their epistemic goals, as discussed in the dimensions listed in Table One, is one good way to accomplish that. Levins's original analysis and follow-on literature provide some justification for developing diverse models in this way, but they do not offer ideas on how to actually engineer models to be different from one another. A key part of our contribution to the Levins literature is our provision of a more detailed taxonomy of independence as well as a motivation for utilizing Levins's notion of independence and for creating independence in practice. Future experimental work could try to identify which of the parameters of independence—characteristics and epistemic goals—matters most when using multiple models and can also assess whether the deliberate engineering of independence into a model group can be useful in practice.

There is of course a tradeoff involved in seeking new and independent models: it takes time and effort to pursue independence, there is never an obvious point to determine when one has enough or the right kind of independence, and there is always a chance that one may model a system and fail to gain insight into it. As a result of these challenges, the answer to whether more independence is worthwhile changes depending on context, and an expert or manager in the given modeling area may be in the best position to assess what level of independence is enough.[35] However, our experience indicates that there are many situations where sufficient resources exist to support a deliberate focus on independence among models, whether aimed towards assessing the amount of independence among models that already exist, revising existing models to be more independent, or even creating new, independent models. Independence may not be as frequently pursued in those ways because of implicit assumptions about what the goals of modeling should be. Getting modeling experts and funders to reflect on what types of independence matter for the system they are addressing could be important and can help provide a way for modelers to talk to stakeholders about uncertainty. Managers and policy-makers can also ask questions of modelers about the different dimensions of independence being used and explore whether there is enough independence. By deepening discussion of relevant types of independence, one can help increase the transparency and accountability of modelers toward outside stakeholders.

Beyond potential practical implications, we have made several theoretical contributions. First, we clarified the literature on Levins's analysis by highlighting independence as core to his approach: increased confidence in the accuracy of model agreement occurs only if models are at least partially independent.[36] Second, the dimensions of independence that we use to show how models can be independent from one another are also new to the literature. Much of the literature has focused on very coarse-grain notions of independence (Odenbaugh and Alexandrova 2011). While we cite Levins's notions of epistemic desiderata and develop them as part of these dimensions of independence, we refined them significantly by breaking out sub-categories of some of the desiderata, changing them based on our own judgment and the case study at hand. Third, we show how independence is qualitative and multi-dimensional. Our case study showed how multiple, competing dimensions of independence came into focus in the levee case study, with some outside groups criticizing the IPET report for a lack of independent social groups among those who were performing the analysis even while the models differed greatly in what and how they represented and modeled the levee system. Further theoretical research in this area could in turn be practically important. While more empirical work needs to be done, we think it likely that a deliberate reflection on modeling strategies would lead some communities to emphasize independence-focused strategies. We show a way in which ‘conceptual engineering,’ in this case of engineering independence into groups of models, can help engineers better serve the needs of society in a variety of areas.

Acknowledgements

We wish to acknowledge two anonymous reviewers from the journal for significant and thoughtful commentary. We also thank Dan Hicks for very insightful comments on a prior draft. Pirtle owes thanks to Jason Scott Robert, Daniel Sarewitz, Braden Allenby, Ryan Meyer, Charles Hunt, Glenn Butts, Jay Falker and Jack Kelly for advice and discussions.

Notes

1. All opinions expressed in this paper are those of the authors and do not necessarily reflect the views of NASA or the United States Government.

2. Due to reasons of space, we will not review critical literature about the role of models in decision making or policy process. Interested readers can refer to Oreskes, Shrader-Frechette, and Belitz (1994), who argue that “the primary value of models is heuristic,” and advocate caution for modelers trying to get involved in aiding policy decisions. While Oreskes, Shrader-Frechette, and Belitz offer some valuable commentary about the impossibility of verification for models of open systems, they do not provide any positive suggestions about the use of models in the policy process. Careful, deliberate reflection on model-building strategy, leading to model groups with more independence or diversity among the individual models, can be one path toward giving positive suggestions on how to model for policy.

3. De Weck, Roos, and Magee say: “Engineering systems attempt to bring these perspectives together into a set of complementary methods and a unified approach that yields a richer set of insights than what could be obtained from each of the classical disciplines alone” (2011, 97). Not many engineering systems monographs offer systematic advice on modeling. In their chapter on modeling, they give advice about how to establish the system boundary of the model, assessing functions (and subfunctions) of the system, defining the structure of the model; and quantifying temporality through simulation. They do not touch on the epistemology literature, including the issues touched on here by Levins, nor on the epistemic goals of models as described here. For a detailed discussion for constructing models in ecology, evolutionary biology, and epidemiology, see (Otto and Day 2007).

4. There are also model-based systems engineering (MBSE) efforts trying to develop a unified language, connecting different functions of a system (Estefan 2008).

We do not examine model-based design in detail in this paper, but some of the same questions about how many (and how independent) MBSE models should be used can still apply.

5. For Breiman, a data modeling strategy focuses on performing regression analysis between inputs and outputs, assuming some type of direct connection between inputs and outputs. An algorithmic modeling strategy can have a much more complex setoff decision rules, reflecting a deeper set of factors. These two cultures are not meant to be exhaustive of all modeling approaches, and are focused on Breiman's field of statistics.

6. We focus in this paper on a strategy of using deliberately independent models to assess the same system, but there is a related phrase that may apply here. Many modelers use the phrase 'model triangulation' or multi-method research to compare model results of the same system from different models. While these are commonly used terms in modeling, Balaban, Hester, and Diallo (2015) note that "the current level of theoretical, methodological and pragmatic knowledge related to a multi-method modeling and simulation approach is limited as there are no clearly identified theoretical principles that guide the use of multi-method M&S approach" (1633) There is relatively little literature discussing model triangulation itself as a concept and exploring what it means, but we note some of that literature in note 28.

7. Recent conceptual treatments of modeling (Weisberg 2013; Winsberg 2010) discuss engineering examples but don't find a meaningful difference between scientific and engineering models. Weisberg focuses in detail on geotechnical engineering models of the San Francisco bay, which he casts as part of a category of concrete models, of which our levees example can be relevant. Pirtle (2010) notes examples where engineering models are made using different types of assumptions and knowledge sources that differ from science but leaves any more meaningful differences between science and engineering as an open question. Given that engineers have practical goals and constraints similar to those that Levins lays out, as discussed later, we think the application of Levins's model pluralism to engineering should face no immediate conceptual barriers.

8. Lloyd's analysis of independence among climate models is valuable, but it largely attempts to summarize epistemic practices used already by climate scientists, and does not try to provide practical advice on how better to perform modeling. The ability to give advice on how to increase independence in groups of models is part of the new contribution being suggested here.

9. Lloyd and Weisberg seem to focus on Levins's discussion of "a common biological assumption" among models (Levins 1966, 423). Lloyd (2015) focuses her reading of robustness on having a "common structure" and "causal core" among models (59), implying that the value of agreement among multiple models requires that they share a causal core. This partially follows Weisberg (2006b), for whom the "com-

mon structure" among models is the shared causal "structure common to the models that gives rise to the robust property" (737) that is agreed upon across models. Our reading of Levins 1966's use of models (see note 15) and ensuing analysis does not make us believe that identifying a common causal core was central to his analysis. Further, we believe that Humphreys (2009), Winsberg (2010), and Odenbaugh and Alexandrova (2011) raise doubts about the achievability of clearly identifying shared causal structures across models. Regardless of the viability of Lloyd's and Weisberg's readings and what Levins's original intent was, we do not make a common causal core to be central to our analysis of the value of robustness and independence, where we focus more on ensuring the use of genuinely independent models. Such a focus can be supported by Levins's claim of having "alternative models each with different simplifications" (Levins 1966, 423). Pirtle (2009) reviews other concepts of independence that can support this approach, such as Hacking 1985 and Wimsatt 2007, which show the value of assessing independence beyond assessing causal structure of a model, such as by increasing likelihood of a correct claim as a result of having more lines of evidence each with separate probabilities of truth.

10. As was reviewed by Odenbaugh (2003, 2006) there has been a long running stream of criticism and defense surrounding the notion of independence as robustness. Soler et al. 2012 have reflected on this, and others criticize the virtue of independence focused approaches (Stegenga 2012 and Orzack and Sober 1993). The general consensus in the academic literature tentatively seems to be in favor of the value of a robustness-as-independence based approach as having some value. Enough if no definitive proof can be made for its value, robustness is invoked commonly enough by scientists that it merits research (Pirtle, Meyer, and Hamilton 2010 illustrate these claims in climate science), and it may have value as a heuristic for discovery the value of which should be assessed empirically for a prospective, new situation (Wimsatt 2007).

11. There is an additional educational reason to discuss a brute force tendency in modeling: Some of the authors, in reflecting on being students and on observing how others learn modeling, know that it can be a very common instinct for new students to 'double down' on adding details into a model. While some students have professors that caution against putting too many details into a model, many intuitively feel that they can overcome complexities by maximizing the fidelity of the model. This also biases them to look to the model for the answer as opposed to other types of analysis that may be less formal. Awareness of alternative epistemic strategies can help mitigate those biases.

12. Brute force modeling may be easier to implement when a system is available to be experimented on, or tested, on a frequent and recurring basis. Automobiles, aircraft and skyscrapers are complex yet engineers have comprehensive, realistic models that can be used to make reliable and safe designs. However, in each case, engineers use experimentation and past experience to validate their design at multiple steps.

The ability to test and intervene on a design prototype is critical to making models of automobiles and aircraft that are accurate and reliable. The ability to experiment is key to making accurate models of complex systems but experimentation is not always possible. For complex sociotechnical engineering systems, it is not always possible to have controlled experiments that can make a brute force modeling strategy successful. It is also not possible in the case examined below, assessing a past engineering failure, to get exact access to the initial conditions of failure, much less to exactly recreate the hurricane and levee conditions that led to the initial failure.

13. We will not attempt to discuss the ontology and realism of models generally, as that is well-explored by many of the references we cite (Weisberg 2013; Winsberg 2010). Levins does have one quote that well captures his views here: “A mathematical model is neither an hypothesis or a theory. Unlike the scientific hypothesis, a model is not verifiable directly by experience. For all models are both true and false. Almost any plausible proposed relation among aspects of nature is likely to be true in the sense that it occurs (although rarely and slightly). Yet all models leave out a lot and are in that sense false, incomplete, inadequate” (Levins 1966, 430).

14. For a related but distinct look at challenges in understanding algorithms based on engaging with their narratives, see Finn 2017.

15. Levins’s 1966 models I–III are not the focus of this paper, and will only briefly be mentioned. The goal of the models was to assess how species evolve in uncertain environments. Two of the models relied upon Levins’s fitness set modeling approach to project how evolution occurs, but they differed significantly in how fitness was calculated. The third model differed in that it had a genetic basis. If there was one common assumption among the three models, it was likely some general theory of evolution, that fitness shapes successive generations. Per endnote 9, it is not clear what a core “causal structure” would be, and his analysis focuses on agreement coming from substantially different models.

16. William Whewell wrote extensively on consilience in the 1800s; while the concept is not in broader use in the philosophy of science today, it is still discussed (Oreskes 2000; Laudan 1971; Wilson 1999). The literature on robustness does not generally reflect on consilience.

17. This tension in our argument was noted by Dan Hicks. We believe that the answer of which strategy is better for a given case will be context dependent, and is best assessed empirically. To do such an assessment of the value of a pluralistic versus monistic versus brute force strategy, there first needs to be more deliberate experimentation in creating and using groups of models that are relevantly independent from one another. The caveats about modeling for policy in Sarewitz, Pielke, and Byerly (2000) and Oreskes, Shrader-Frechette, and Belitz (1994) likely do imply that simplicity can have major inherent advantages when dealing with complex, value-laden problems.

18. There was no explicit debate in the report about using something akin to a 'brute force' versus a 'pluralistic' modeling strategy. It was simply assumed that it was good practice to use multiple models, with the implicit claim that increased confidence would arise from using the combination. In the sense of an 'ad hoc' or 'implicit' modeling strategy noted above, it seems like the geotechnical engineering field may be somewhere in between the two.

19. We thank Springer Press for permission to use part of this discussion of levees, from Pirtle 2010. It has been heavily elaborated and modified here.

20. These and some of the other explicit details about the FEA/LEA/Centrifuge models comes from the IPET detailed technical appendices located in IPET 2007 Appendixes 4–6.

21. Wimsatt and others' expansion of Levins's framework does focus on getting increasingly independent forms of analysis, making the desire for more independence expand beyond focusing on numerical models. Baird's argument that shows how physical models can carry knowledge also puts them on equal par with numerical models. Weisberg 2013 also discusses 'concrete,' physical models as being its own type of model category alongside mathematical and numerical models, which provides further context for treating the centrifuge model here as part of a Levins-like group of models that can support a shared claim.

22. Subsequent literature notes some disagreements with the IPET approach but often is confirming: Adhikari, Song, and Cheng 2014 came to different conclusions than the IPET on some of the soil strength and moduli but argues this did not "affect the failure mechanism of the levee substantially" (1123). The IPET 2009 discussion of Bea's results are on V-125/6, which says "The UCB [Bea] Team hypothesized that a thin 1-in.-thick sensitive clay layer within the [marsh] peat layer (overlying the weak clay) is the location of the slip plane that caused the failure. The UCB Team based this on . . . soil samples and limited numerical analyses using a suspect geology profile and unrealistic soil properties. . . . IPET has not been able to detect a widespread thin (inch or less in thickness) layer within the peat layer, in spite of having physically harvested large volumes of peat for use in centrifuge testing and acquisition of significant additional subsurface soils data." If the weak soil had been higher, then it implies that it should have been more easily detected and reflects negligence. Bea's team did agree with the IPET team about a type of gap being formed before failure.

23. How the models were built to reach this point would be a richer and more complex story. The LEA, FEA and centrifuge model concepts all existed before the Katrina failure, and in a key-sense the IPET's model strategy was about the 'usage' of models and not the 'building' of models.

24. Efforts to combine or align the models could have been possible, such as by attempting to combine FEA and LEA modeling approaches (though it is not clear that such an approach is possible), or to more closely tie the FEA/LEA models to the

centrifuge. One could even imagine a brute force but hopeless model that attempts to simulate every atom at the 17th St. Canal during the failure, including both the ground and each successive wave of water. Such brute force approaches to any problem will suffer to due uncertainty in data, much less being able to meaningfully model interconnections at an atomic level across a large scale.

25. If one is modeling a system that is complex and where a brute force modeling approach seems unlikely to work, then implementing a pluralistic modeling approach may be necessary in some contexts, and may be desirable in most. Paying close attention to independence among models is vital to that approach. Considering these dimensions can allow one to better understand what the existing model-building strategy is in a field of engineering, and in turn to contrast the model strategy of that discipline to other fields. By clearly identifying and reflecting on what a field's current model strategy is, one can be better prepared to reflect on what the model-building strategy should be.

26. Giere (1988, chap. 3) and Nelson Goodman (1970) make this point.

27. Statistical independence involves assessing whether two events have the same likelihood of occurrence—whether the occurrence of one entails a likelihood of the other happening. This concept has less applicability to modeling than does a related definition, where model independence can be defined based on whether two models are likely to give the same output based upon having the same input. See the discussion of Abramowitz and Gupta 2008 in Pirtle, Meyer, and Hamilton 2010. The definitions discussed below are more about the inputs and nature of the models than about an output-based comparison approach. Both are important, and the notions of independence discussed here can deepen a discussion of the independence of model outputs.

28. An indirectly related concept to independence is the concept of model triangulation. Balaban, Hester, and Diallo (2015) discusses some of the history of triangulation, including how the concept has been applied in debates about what metamethodologies are. Denzin (1970) discusses different types of triangulation such as triangulation across different datasets, across different people assessing an object, and triangulation across use of different methods. The physical and social categories of independence that we discuss later are similar and more nuanced than these categories, but the epistemic desiderata that we discuss are significantly different.

29. I am elaborating on these dimensions in Pirtle (In preparation), where additional description of these dimensions of independence can be found.

30. An eloquent description of three different levels of abstractions among models can be found in Gilbert 2008 (section 3.3): he refers to abstract models, which may not bear relations to real systems except in an abstract, theoretical way; 'middle range' models, which describe a phenomenon but in a way that it can be generally extended to multiple empirical systems; and facsimile models, which are meant to be a close

reproduction of a real world system or phenomenon. The engineering systems scholar Michael Pennock has been studying the practice of multi-level modeling which uses models of different levels of abstraction to assess a given system, and hopes to assess best practices for combining results from multiple levels of a system (Pennock and Gaffney 2016).

31. Levins continues the above quote: "But this cannot be done. Therefore, several alternative strategies have developed." Simply put, Levins claims that is impossible to have "models which maximize generality, realism and precision" (Levins 1966, 422). Although Levins never used the phrase in 'The Strategy,' many authors refer to the following point as the 'tradeoff hypothesis' (Weisberg 2006a). In 'The Strategy,' Levins does not explicitly define realism, generality and precision, which led to criticisms by Orzack and Sober (1993) that Levins does not explicitly justify the existence of tradeoff. The objections made in Orzack and Sober 1993 have been reviewed and mostly rejected in Levins 1993, Odenbaugh 2003, Odenbaugh 2006, Weisberg 2006a, and Weisberg 2006b. As Levins 1993 acknowledges that science often has an evolving and progressive state of the art that can make progress overtime, his claim about tension between epistemic desiderata is more about the experience of an individual modeler. Given limited time and effort, it can be difficult for a modeler to have a perfect model along each of these desiderata. Levins's hypothesis would recommend picking one or two main epistemic goals at a time. This claim also applies generally to the topic of brute force modeling, as the challenges experienced there can change and ebb over time.

32. Per the hypothetical in note 24, even a model that tries to examine each atom of a system would have to make assumptions: does it show the space between the parts of the atom, and how does it deal with describing the location of each of those pieces? Adding in more details requires the use of more assumptions about how the details relate, which can require arbitrary, uncertain or false assumptions to incorporate them.

33. An anonymous reviewer asked about independence if the same modeler is creating two different models. Per the list in Table One, an individual modeler can definitely create models that are independent to one another in what their causal mechanisms are, how they represent the system, and in terms of how general, realistic and precise they are meant to be, etc. But, of course, those models are likely being strongly influenced by concepts that the individual might hold—their training, personal biases and other dimensions which may strongly influence the overall results. That influence and perception may also lead them to think there is more independence among the other dimensions listed above than there really are. For truly important problems, one wants to have multiple groups assess the problem, with consideration given to conflicts of interest and having diverse skillsets being used. On the other hand, it is possible for independent outside groups to assess a given system while having no meaningful

independence among the more ‘technical’ dimensions of the model. For this reason, discussions of independence should include both social and technical factors.

34. An intentional ‘strategy’ involves awareness of how models differ from one another, including how they variously reflect different model goals such as generality, realism, precision and manageability. We contend that a mature strategy of model building should consider carefully the complexity of the system, the limits of data available, and have a proactive plan for assessing what types of models and modelers should be used, how different they should be, and how much they should differ on criteria such as precision and generality.

35. A reviewer asked what happens if there is a disagreement on whether more independence is needed amongst a group of models: one analyst might think more diverse viewpoints are needed to assess the problem, and another might view it as a straightforward problem that is sufficiently assessed with one model. This is a difficult point to answer: experts can disagree on any topic in modeling, and there is rarely a way to get an objective answer to a problem outside of continued peer review, empirical research and discussion in the literature. By shifting the debate to the different dimensions of independence discussed above, it is possible that debate about whether more independence is needed can be clarified and there can be an easier way to get agreement among analysts, as well as to elicit input and explain the debate to other stakeholders such as elected or appointed decision-makers. However, it is possible that there simply may not be agreement on the need or lack-thereof for an independence-focused strategy. This is not surprising, as we have already noted that an independence-focused framework can be fallible. Eventually, if there is a pattern of failure or a lack of progress on assessing a given topic, analysts will have more perspective with which to assess the results and a more objective answer on whether independence is needed might be obtained.

36. In the parlance of Levins and other literature that follows him, this would say that ‘robustness’ benefits only accrue if the input models are at least partially independent. As mentioned before, we have avoided the word ‘robustness’ in this paper because the word has deeply entrenched meanings in other fields (often tied to assessing system features that are ‘robust’ to parameter variation), which can cause confusion.

References

Abbott, Russ. 2007. “Putting Complex Systems to Work.” *Complexity* 13(2): 30–49. https://doi.org/10.1002/cplx.20200

Abramowitz, Gab, and Hoshin Gupta. 2008. “Toward a Model Space and Model Indpendence Metric.” *Geophysical Research Letters* 35(5): 4. https://doi.org/10.1029/2007GL032834

Adhikari, Sudarshan, Chung R. Song, and Alexander H.-D. Cheng. 2014. "Evaluation of I-wall in New Orleans with Back-Calculated Total Stress Soil Parameters." *Acta Geotechnica* 9: 1123–34. https://doi.org/10.1007/s11440-013-0264-1

Argyropoulos, C. D., and N. C. Markatos. 2015. "Recent Advances on the Numerical Modelling of Turbulent Flows." *Applied Mathematical Modelling* 39(2): 693–732. https://doi.org/10.1016/j.apm.2014.07.001

Baird, Davis. 2004. *Thing Knowledge: A Philosophy of Scientific Instruments*. Berkeley: University of California Press.

Balaban, Mariusz, Patrick T. Hester, and Saikou Diallo. 2015. "Towards a Theory of Multi-Method M&S Approach: Part III." In *Proceedings of the 2015 Winter Simulation Conference*, ed. L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, 1633–44. New York: Institute of Electrical and Electronics Engineers. https://doi.org/10.1109/WSC.2015.7408283

Breiman, Leo. 2001. "Statistical Modeling: The Two Cultures (with Comments and a Rejoinder by the Author)." *Statistical Science* 16(3): 199–231. https://doi.org/10.1214/ss/1009213726

Denzin, Norman K. 1970. *The Research Act in Sociology: A Theoretical Introduction to Sociological Methods*. Chicago: Aldine.

Dessai, Suraje, Mike Hulme, Robert Lempert, and Roger Pielke Jr. 2009. "Do We Need Better Predictions to Adapt to a Changing Climate?" *Eos Transactions, American Geophysical Union* 90(13): 111–12. https://doi.org/10.1029/2009EO130003

De Weck, Olivier L., Daniel Roos, and Christopher L. Magee. 2011. *Engineering Systems: Meeting Human Needs in a Complex Technological World*. Cambridge, MA: MIT Press.

Duncan, J. Michael, and Stephen G. Wright. 2005. *Soil Strength and Slope Stability*. Hoboken, NJ: John Wiley and Sons.

Edwards, Paul N. 2010. *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming*. Cambridge, MA: MIT Press.

Estefan, Jeff A. 2008. "Survey of Model-Based Systems Engineering (MBSE) Methodologies." *Incose MBSE Initiative*. http://www.omgsysml.org/MBSE_Methodology_Survey_RevB.pdf.

Finn, Ed, 2017. *What Algorithms Want: Imagination in the Age of Computing*. Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/9780262035927.001.0001

Giere, Ronald N. 1988. *Explaining Science: A Cognitive Approach*. Chicago: University of Chicago Press. https://doi.org/10.7208/chicago/9780226292038.001.0001

Giere, Ronald N. 1999. *Science without Laws*. Chicago: University of Chicago Press.

Gilbert, Nigel, 2008. *Agent-Based Models*. Los Angeles: Sage Publications. https://doi.org/10.4135/9781412983259

Goodman, Nelson. 1970. "Seven Strictures on Similarity." In *Experience & Theory*, ed. Lawrence Foster and Joe W. Swanson. Amherst: University of Massachusetts Press.

Hacking, Ian. 1985. "Do We See through Microscopes?" In *Images of Science: Essays on Realism and Empiricism*, ed. Paul Churchland and Clifford. A. Hooker, 132–52. Chicago: University of Chicago Press.

Hamaker, Joseph W. 2010. "Do More Details Imply Better Accuracy in Cost Estimates?" *Parametric World* 29(3): 13, 22. Available at http://www.iceaaonline.com/ready/wp-content/uploads/2014/03/2010-Parametric-World-Summer.pdf.

Humphreys, Paul. 2009. "The Philosophical Novelty of Computer Simulation Methods." *Synthese* 169(3): 615–26. https://doi.org/10.1007/s11229-008-9435-2

Interagency Performance Evaluation Taskforce. 2007. *Performance Evaluation of the New Orleans and Southeast Louisiana Hurricane Protection System, Final Report of the Interagency Performance Evaluation Task Force. Volume V—The Performance—Levees and Floodwalls*. https://ipet.wes.army.mil/.

Kestenbaum, David. 2006. "Why Did the 17th St Canal Levee Fail?" *All Things Considered*. National Public Radio, May 19.

Knorr Cetina, Karin. 2009. *Epistemic Cultures: How the Sciences Make Knowledge*. Cambridge, MA: Harvard University Press.

Knutti, Reto. 2018. "Climate Model Confirmation: From Philosophy to Predicting Climate in the Real World." In *Climate Modeling: Philosophical and Conceptual Issues*, ed. Elizabeth A. Lloyd and Eric Winsberg, 325–59. Dordrecht: Springer.

Laudan, Larry. 1971. "William Whewell on the Consilience of Inductions." *The Monist* 55(3): 368–91. https://doi.org/10.5840/monist197155318

Levins, Richard. 1966. "The Strategy of Model Building in Population Biology." *American Scientist* 54(4): 421–31.

Levins, Richard. 1968. "Ecological Engineering: Theory and Technology; A Review of *Ecology and Resource Management: A Quantitative Approach* by Kenneth Watt." *The Quarterly Review of Biology* 43(3): 301–05. https://doi.org/10.1086/405813

Levins, Richard. 1993. "Reply to Orzack and Sober." *Quarterly Review of Biology* 68(4): 547–55. https://doi.org/10.1086/418302

Lloyd, Elizabeth A. 2015. "Model Robustness as a Confirmatory Virtue: The Case of Climate Science." *Studies in History and Philosophy of Science Part A* 49: 58–68. https://doi.org/10.1016/j.shpsa.2014.12.002

Masson, D., and Reto Knutti. 2011. "Climate Model Genealogy." *Geophysical Research Letters* 38(8). https://doi.org/10.1029/2011GL046864

Odenbaugh, Jay. 2003. "Complex Systems, Trade-offs, and Theoretical Biology: Richard Levins's 'The Strategy of Model Building in Population Biology' Revisited." *Philosophy of Science* 70(5): 1496–1507. https://doi.org/10.1086/377425

Odenbaugh, Jay. 2006. "The Strategy of 'The Strategy of Model Building in Population Biology.'" *Biology and Philosophy* 21(5): 607–21. https://doi.org/10.1007/s10539-006-9049-3

Odenbaugh, Jay, and Anna Alexandrova. 2011. "Buyer Beware: Robustness Analyses in Economics and Biology." *Biology & Philosophy* 26(5): 575–71. https://doi.org/10.1007/s10539-011-9278-y

Oreskes, Naomi, 2000. "Why Predict? Historical Perspectives on Prediction in Earth Science." In *Prediction: Science, Decision Making, and the Future of Nature*, ed. Daniel Sarewitz, Roger A. Pielke Jr., and Radford Byerly Jr., 23–40. Washington, DC: Island Press.

Oreskes, Naomi, Kristin Shrader-Frechette, and Kenneth Belitz. 1994. "Verification, Validation, and Confirmation of Numerical Models in the Earth Sciences." *Science* 263(5147): 641–46. https://doi.org/10.1126/science.263.5147.641

Orzack, Steven H., and Elliot Sober. 1993. "A Critical Assessment of Levins's 'The Strategy of Model Building in Population Biology' (1966)." *Quarterly Review of Biology* 68: 533–46.

Otto, Sarah. P., and Troy Day. 2007. *A Biologist's Guide to Mathematical Modeling in Ecology and Evolution*. Princeton, NJ: Princeton University Press.

Parker, Wendy S. 2011. "When Climate Models Agree: The Significance of Robust Model Predictions." *Philosophy of Science* 78(4): 579–600. https://doi.org/10.1086/661566

Parker, Wendy S. 2018. "The Significance of Robust Climate Projections." In *Climate Modeling: Philosophical and Conceptual Issues*, ed. Elizabeth A. Lloyd and Eric Winsberg, 273–96. Dordrecht: Springer. https://doi.org/10.1007/978-3-319-65058-6_9

Pennock, Michael J., and Christopher Gaffney. 2016. "Managing Epistemic Uncertainty for Multimodels of Sociotechnical Systems for Decision Support." *IEEE Systems Journal* 12(1): 184–95. https://doi.org/10.1109/JSYST.2016.2598062

Pirtle, Zachary. 2009. "Robustness Analysis: Measuring Independence within Multi-Model Ensembles." Thesis for the MS degree in Civil and Environmental Engineering. Arizona State University.

Pirtle, Zachary. 2010. "How the Models of Engineering Tell the Truth." In *Philosophy and Engineering: An Emerging Agenda*, ed. Ibo van de Poel and David Goldberg, 95–108. Dordrecht: Springer.

Pirtle, Zachary. In preparation. "Engineering Independence: An Analysis Framework for Systems Engineering in Complex Development Programs." Dissertation, Department of Engineering Management and Systems Engineering, The George Washington University.

Pirtle, Zachary, Ryan Meyer, and Andrew Hamilton. 2010. "What Does It Mean When Climate Models Agree? A Case for Assessing Independence among General Circulation Models." *Environmental Science and Policy* 13(5): 351–61. https://doi.org/10.1016/j.envsci.2010.04.004

Puccia, Charles, and Richard Levins. 1985. *Qualitative Modeling of Complex Systems*. Cambridge, MA: Harvard University Press. https://doi.org/10.4159/harvard.9780674435070

Sarewitz, Daniel, Roger A. Pielke Jr., and Radford Byerly, eds. 2000. *Prediction: Science, Decision Making, and the Future of Nature*. Washington, DC: Island Press.

Seed, R. B., R. G. Bea, R. I. Abdelmalak, A. G. Athanasopoulos, G. P. Boutwell, J. D. Bray, J.-L. Briaud, C. Cheung, D. Cobos-Roa, J. Cohen-Waeber, B. D. Collins, L. Ehrensing, D. Farber, M. Hanemann, L. F. Harder, K. S. Inkabi, A. M. Kammerer, D. Karadeniz, R. E. Kayen, R. E. S. Moss, J. Nicks, S. Nimmala, J. M. Pestana, J. Porter, K. Rhee, M. F. Riemer, K. Roberts, J. D. Rogers, R. Storesund, A. V. Govindasamy, X. Vera-Grunauer, J. E. Wartman, C. M. Watkins, E. Wenk Jr., and S. C. Yim. 2006. *Investigation of the Performance of the New Orleans Flood Protection System in Hurricane Katrina on August 29, 2005. Volume I: Main Text and Executive Summary*. http://projects.ce.berkeley.edu/neworleans/report/VOL_1.pdf.

Shackley, Simon, Peter Young, Stuart Parkinson, and Brian Wynne. 1998. "Uncertainty, Complexity and Concepts of Good Science in Climate Change Modelling: Are GCMs the Best Tools?" *Climatic Change* 38(2): 159–205. https://doi.org/10.1023/A:1005310109968

Soler, Léna, Emiliano Trizio, Thomas Nickles, and William Wimsatt, eds. 2012. *Characterizing the Robustness of Science: After the Practice Turn in Philosophy of Science*. Dordrecht: Springer. https://doi.org/10.1007/978-94-007-2759-5

Stegenga, Jacob. 2012. "Rerum concordia discors: Robustness and Discordant Multimodal Evidence." In *Characterizing the Robustness of Science: After the Practice Turn in Philosophy of Science*, ed. Léna Soler, Emiliano Trizio, Thomas Nickles, and William Wimsatt, 207–26. Dordrecht: Springer. https://doi.org/10.1007/978-94-007-2759-5_9

Vezér, Martin A. 2016. "Computer Models and the Evidence of Anthropogenic Climate Change: An Epistemology of Variety-of-Evidence Inferences and Robustness Analysis." *Studies in History and Philosophy of Science* 56: 95–102. https://doi.org/10.1016/j.shpsa.2016.01.004

Waters, Kenneth. 2004. "What Concept Analysis in Philosophy of Science Should Be (and Why Competing Philosophical Analyses of Gene Concepts Cannot Be Tested by Polling Scientists)." *History and Philosophy of the Life Sciences* 26(1): 29–58. https://doi.org/10.1080/03919710412331341631

Weisberg, Michael. 2006a. “Forty Years of ‘The Strategy’: Levins on Model Building and Idealization.” *Biology and Philosophy* 21(5): 623–45. https://doi.org/10.1007/s10539-006-9051-9

Weisberg, Michael. 2006b. “Robustness Analysis.” *Philosophy of Science* 73(5): 730–42. https://doi.org/10.1086/518628

Weisberg, Michael. 2013. *Simulation and Similarity: Using Models to Understand the World*. New York: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199933662.001.0001

Wilson, Edward O. 1999. *Consilience: The Unity of Knowledge*. New York: Vintage.

Wimsatt, William. 2007. *Re-Engineering Philosophy for Limited Beings: Piecewise Approximations to Reality*. Cambridge, MA: Harvard University Press.

Winsberg, Eric. 2010. *Science in the Age of Computer Simulation*. Chicago: University of Chicago Press. https://doi.org/10.7208/chicago/9780226902050.001.0001

Winsberg, Eric. 2012. “Values and Uncertainties in the Prediction of Global Climate Models.” *Kennedy Institute of Ethics Journal* 22(2): 111–37. https://doi.org/10.1353/ken.2012.0008