

The Elusive Basis of Inferential Robustness

James Justus*†

Robustness concepts are often invoked to manage two obstacles confronting models of ecological systems: complexity and uncertainty. The intuitive idea is that any result derived from many idealized but credible models is thereby made more reliable or is better confirmed. An appropriate basis for this inference has proven elusive. Here, several representations of robustness analysis are vetted, paying particular attention to complex models of ecosystems and the global climate. The claim that robustness is itself confirmatory because robustness analysis employs a Bayesian variety-of-evidence argument is criticized, but recent overwhelming pessimism about robustness may have a silver lining.

1. Introduction. Attempts to model ecological systems and make reliable decisions about environmental issues face two related and formidable obstacles: complexity and uncertainty. The complexity of ecosystems often precludes representing them with high fidelity. Doing so would require model variables, parameters, and equations that capture every nonnegligible feature of a target system's structure and dynamics. For simple systems, this is sometimes feasible; for complex systems, it almost never is. To make models of the latter type tractable, simplifying idealizations are therefore necessary. But the typically vast quantities of data needed to assess which idealizations do and do not prevent adequate representation of complex systems are rarely available given the monetary, technological, and temporal limitations of most scientific research, especially in ecological and environmental science. This uncertainty limits what insights the models can deliver about real-world ecosystems and, thus, the degree to which such models provide legitimate factual bases for environmental decision making. As ecologists, policy makers,

*To contact the author, please write to: Philosophy Department, Florida State University, Tallahassee, FL 32306; e-mail: jjustus@fsu.edu.

†For helpful feedback, thanks to John Matthewson, Carl Salk, Jan Sprenger, and audiences at the Philosophy of Science Association Biennial Meeting at Montreal (November 2010) and the International Society for the History, Philosophy, and Social Studies of Biology at Salt Lake City (July 2011). Special thanks to Jan Sprenger for organizing the PSA session in which this article was presented.

Philosophy of Science, 79 (December 2012) pp. 795–807. 0031-8248/2012/7905-0025\$10.00
Copyright 2012 by the Philosophy of Science Association. All rights reserved.

and environmentalists appreciate, or should, complexity and uncertainty make ecological modeling and environmental decision making exceedingly difficult.

Analyses of problems involving such high degrees of complexity and uncertainty often appeal to robustness concepts. Informally, robustness analysis is a kind of inferential bootstrapping. A set of data sources, estimation methods, experimental designs, or models is identified. Each set member is taken to triangulate the truth. For models, set members are treated as *plausible* representations of the real-world system. But plausibility is largely gauged in an inexact way on the basis of past predictive success and the presumption that model assumptions reasonably reflect reality; the precise representational fidelity is unknown. The expectation is that any result derived from many different models, each making credible but incompatible assumptions about the real-world system being represented, is thereby made more reliable or is better confirmed. Robustness analysis could therefore be a basic tool of inference in ecology, environmental science, and other sciences of complex systems.

As an aspiring general method of scientific inference, robustness analysis should be representable (and vindicated) within general statistical philosophies, such as Bayesianism. Where different data sources are the focus, the legitimacy of some types of robustness analysis is well established (see Franklin and Howson 1984; Earman 1992; Fitelson 2001). But finding defensible characterizations of other types has proven difficult, particularly where robustness involves a plurality of different models. For example, formulated simplistically as an inference method in which a robust result follows deductively from every model, robustness analysis has very limited applicability and thus offers little inferentially (Orzack and Sober 1993). Incompatible conclusions about the value of robustness have also been reached for more realistic, and often probabilistic inferences. In a study of Lotka-Volterra predator-prey models, Weisberg (2006) defends and elaborates the scientific utility of robustness analysis but argues that robustness itself is not confirmatory. In a recent analysis of climate models, however, Lloyd (2009, 2010) argues that robustness increases confirmation.

One limitation of previous studies is that general statistical philosophies that ground scientific inference receive little consultation concerning the new robustness candidate. This shortcoming is addressed by analyzing a recent Bayesian representation of robustness analysis. Particular attention is paid to whether it can play the role it is accorded in ecology and environmental science, and climate modeling specifically. Section 2 begins with Richard Levins's seminal account of robustness, Orzack and Sober's (1993) criticism, and Levins's (1993) little-considered response. Section 3 describes Weisberg's (2006) rehabilitation of robustness analysis, but section 4 argues that complex systems still pose significant challenges. Global climate mod-

eling, in particular, reveals important limitations of the methodology. Section 5 criticizes the claim that Weisberg's robustness analysis is a Bayesian variety-of-evidence inference and thereby shows robustness is itself confirmatory. Section 6 briefly concludes that, despite the negative assessments of this and most other analyses, there are reasons to believe that a cogent basis for the confirmatory value of robustness will eventually be developed.

2. Truth as the Intersection of Independent Lies. Idealizations that make simplifying, unrealistic assumptions for the sake of tractability are a kind of lie about the target real-world system being represented. But the individual distortions of different models may belie a common veridicality that supplies reliable insights. This was the basis of Levins's robustness methodology:¹ "we 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" (1966, 423). The idea seemed compelling, and its application to different models of niche dynamics was influential, but as an inference method several aspects remained unclear: How many models are "several"? Is it the simplifications (lies) that must be independent or the models themselves, as Orzack and Sober (1993) later assumed? If the latter, how can the models possess a common assumption? What precisely is meant by "independent"? And how are simplifications delimited from the "common biological assumption"? And since almost any claim follows from numerous—perhaps an infinite number of—models, much more guidance was needed about how different and how plausible alternative models must be for the robustness inference to be reliable.

This imprecision prompted Orzack and Sober (1993) to reconstruct and then criticize the proposed inference method. Let R be a so-called robust theorem, that is, a consequence of each of several distinct models M_1, M_2, \dots, M_n intended to represent a particular real-world system. Three formulations were offered:

- i) If one model is known to be true without knowledge of which one, R must be true. This inference is unproblematic but also largely moot. Justified confidence that any finite set of models nets the truth is very rare, especially for complex systems.

1. Wimsatt (1981/2007) generalized the methodology to other means of scientific inquiry, such as estimation procedures and experimental designs. Note that inferential robustness is the focus here, not robustness of the phenomena being inferred about. Notions of the latter may facilitate the former, but they can diverge (see Trout 1993).

- ii) If all the models are known to be false, it is implausible that R acquires any support by being robust. Collective falsehood does not confer support.
- iii) If it is unknown whether any model is true, R seems similarly uncertain. Model uncertainty seems to beget uncertainty, not support for model consequences. If the models exhaustively sampled all possible representations of the target system, R could be inferred. But this antecedent is at least as implausible as the highly implausible antecedent of i.

Given these unappealing options, Orzack and Sober suggested that robustness may simply reflect properties models share rather than anything about the real-world systems they are intended to represent. If so, robustness would poorly guide inference.

His response received little attention, but Levins (1993) argued that this criticism misconstrues his inference methodology by overlooking the differential plausibility of model parts. To make that case, Levins construed each of M_1, M_2, \dots, M_n as an intersection of a common, plausible core C that all the models share and an unshared part V_i unique to model M_i ; V_i represents the simplifying assumptions that M_i makes. Establishing a connection between C and R via M_1, M_2, \dots, M_n is the goal. Robustness analysis then became a two-step procedure:

$$\begin{array}{l}
 \text{(A)} \\
 \text{If } C \cap V_1 \rightarrow R \\
 \quad C \cap V_2 \rightarrow R \\
 \quad \quad \vdots \\
 \quad C \cap V_n \rightarrow R, \\
 \text{then } [C \cap \cup_{i \in I} V_i] \rightarrow R.
 \end{array}$$

$$\text{(B) If } \{V_1, V_2, \dots, V_k\} \text{ "exhausts all admissible alternatives," then } \cup_{i \in I} V_i = 1 \text{ and } C \rightarrow R \text{ (Levins 1993, 553).}$$

Levins suggested inference A is made stronger, and B's antecedent more plausible, as the number of models considered increases. He also stressed that not just any C or V_i will do: C must be plausible, and the V_i reasonable, as gauged by prior observations. With this safeguard, Levins (1993, 554) suggested Orzack and Sober's criticism—that robustness is problematically being invoked as a nonempirical confirmation method, a "way to truth independent of observation"—targets a more ambitious mark than was intended.

This account improves on formulations i–iii in many respects, but serious difficulties remain. First, since the models share C , the model independence

that featured prominently and seemingly indispensably in the original account has been abandoned. Second, Levins does not formulate a criterion, and offers very little specific guidance, about how to delineate C and V . Third, if robustness is not confirmatory, the claim that it can “strengthen the conclusion [R]” (Levins 1993, 554) needs clarification and defense. For these reasons, Levins’s amended account is ultimately inadequate. But the account merits close scrutiny because it contains the seeds of a more compelling version of robustness analysis due to Weisberg (2006), one recently claimed to show that robustness itself is confirmatory in the context of climate modeling (Lloyd 2009, 2010).

3. Rehabilitating Robustness Analysis. Weisberg (2006) presents robustness analysis as a four-step procedure:

1. determine whether a robust property (R) follows from each of M_1, M_2, \dots, M_n ;
2. determine whether M_1, M_2, \dots, M_n share a common structure C ;
3. formulate a ‘robust theorem’ connecting C and R and empirically interpreting it;²
4. assess the scope and strength of the connection with stability analysis.

One of many merits of the analysis is the very well chosen and carefully dissected examples that clarify how robust properties and common structures can be identified, how robust theorems can be interpreted and evaluated, and why such inquiry is highly context dependent. Weisberg examines three simple Lotka-Volterra predator-prey models that make distinct idealizations. A robust property exists that, interpreted ecologically, says a general pesticide would increase prey populations relative to predator populations. Analysis also reveals the common mathematical structure responsible for this robust property that, interpreted ecologically, says prey growth rate primarily controls predator abundance and predator death rate primarily controls prey abundance. Only with intimate knowledge of the particularities of these models could this R and C be identified. Robustness analysis is a case-by-case process; general procedures for finding R and C should not be expected.

Weisberg’s account also recognizes and uses dependencies between models rather than implausibly requiring independence. And in part because the independence assumption has been abandoned, robustness’ epistemic payoff is clear: robust theorems link the empirical support of C and R . In particular, the link establishes two directions for empirical support to flow.

2. Note that ‘robust theorem’ has different designations for Levins and Weisberg.

Case 1: $C \rightarrow R$.

If C were highly supported but R uncertain, support would propagate from the former to the latter via the robust theorem. This is the typical scenario for robustness analysis. High credence in models based on their prior predictive success—and antecedent plausibility given what is known about the systems they represent—is taken to support their core commitments and, in turn, confer support on their joint predictions. On this basis, joint predications of global climate models are sometimes claimed to possess enhanced empirical support (see Parker 2011).

Case 1 was the main focus, but Weisberg briefly alludes to the converse possibility.

Case 2: $R \rightarrow C$.

“If a sufficiently heterogeneous set of models for a phenomenon all have the common structure, then it is very likely that the real-world phenomenon has a corresponding causal structure. This would allow us to infer that when we observe the robust property in a real system, then it is likely that the core structure is present and that it is giving rise to the property” (Weisberg 2006, 739). In this case, empirical support flows from the observationally confirmed robust property R to C . The qualifier “sufficiently heterogeneous” helps guard against the possibility that another structure found outside the set of models considered generates R . But absent guidance about the level of sufficiency and notion of heterogeneousness required, the worry that the models sample from a region of possibility space far from actuality cannot be disregarded. Case 2, not case 1, is the kind of robustness inference Lloyd considers for climate models.

It should be stressed that on this account robustness is “not itself a confirmation procedure” (Weisberg 2006, 732). That R is robust relative to a common core structure C of M_1, M_2, \dots, M_n does not thereby confirm R .³ Rather, robust theorems establish conduits through which empirical support for C can transmit to R , and vice versa. If the C of a robust theorem is highly

3. However, Weisberg (2006) does state that robustness analysis presupposes, but does not supply, a “low-level confirmation” of a mathematical framework that “licenses us to regard the mathematical dependence of [structure C] on [property R] as a causal dependence.” A mathematical framework acquires this confirmation if it “*can* adequately represent the phenomena of interest” (741; emphasis added). Apart from the unusual idea that mathematics itself is confirmable (Sober 1993), however low level, it is unclear whether this modality can do that work. First, it seems highly implausible that richer, more expressive mathematical frameworks are somehow confirmed by their greater representational capability. What matters for confirmation is whether an empirically interpreted mathematical structure *does* adequately represent, not whether the mathematical framework in

confirmed, robustness analysis can establish that the relevant R is thereby highly confirmed. So construed, robustness analysis has a rightful claim as a method or tool of confirmation (cf. Odenbaugh and Alexandrova 2011), even if robustness is not itself confirmatory.

4. The Complexity Challenge Redux. Unfortunately, even this account of robustness analysis has serious limitations for the types of complex system models often found in ecology and environmental science. Examining Lloyd's (2009, 2010) application to climate modeling reveals some of the difficulties.

A brief précis on climate modeling provides helpful background. Global climate models (GCMs) are impressively complicated dynamic system models of processes driving climate at global spatial scales and temporal scales often measured in decades or centuries (Parker 2006). They contain variables, parameters, and equations relating them that number in the hundreds (or more) and form highly complex feedback loops. Despite their individual and collective complexity, significant commonality exists across GCMs. At their foundation are well-established theories of mechanical, fluid, and thermodynamics. They all integrate atmospheric and oceanic dynamics together with the dispersion of solar radiation. And they all involve simulation indispensably: basic equations representing and integrating these dynamics are approximated and solved by simulation.

Apart from commonalities, there are also significant differences. Different GCMs employ different mathematical representations of the atmosphere—for example, the atmosphere as a gridded collection of volumes versus as a series of climatic waves—and different numerical solution techniques. Their empirical assumptions also often diverge. Distinct models contain different parameters, parameter values, and functional relationships between climate drivers that reflect uncertainty about climatic processes. Robustness analysis seems to offer a promising approach to managing this uncertainty.

One variety of robustness analysis with a sound basis and successful track record managing uncertainty is sensitivity analysis. The aim is to evaluate the sensitivity of predictions and properties to specific parameter values, different parameters, and model structures that span the extent of our uncertainty about the target system being represented. If a model prediction or property is largely unaffected by these factors, it is often labeled robust. In this way, robustness analysis as sensitivity analysis helps identify strong and

which it is expressed *can*. Second, confirmation increases probability for almost all theories of confirmation. If mathematical frameworks possess low-level confirmation, this would entail the odd result that empirically interpreted but thoroughly empirically inadequate mathematical structures would nevertheless receive a probability boost from the representationally adept framework they are expressed within.

weak determinants of model predictions and properties as well as dependencies among those determinants. So-called multimodel ensemble methods in recent climate modeling implement sensitivity robustness: different models within the ensemble embodying different assumptions about what drives climate dynamics (see Parker 2010). Averaging predictions from different GCMs to help mitigate potential individual biases is another unassailable tactic this modeling strategy employs. But Lloyd envisages a more ambitious role for robustness considerations in climate modeling.

Lloyd (2009, 220) applies Weisberg's robustness analysis to climate modeling as follows: "we find that in all [GCMs] there is a significant role played by greenhouse gases in the late twentieth-century warming of the global climate, and that these are linked to the surface temperature rising in the equations. . . . Thus, we would have an analysis isolating greenhouse gases linked to temperature rise (the common structure), and a robust theorem linking greenhouse gases to the robust property, the outcome of rising global mean temperature." This is a case 2 ($R \rightarrow C$) application in which C is the relationship between (increases in) greenhouse gases and temperature rise, and R is the robust prediction of a 0.5° – 0.7° C mean global temperature increase in the twentieth century that observations have borne out. Lloyd (2009, 6ff.) states that there are numerous other observationally verified robust joint predictions of GCMs. The claim is that these verified robust predictions redound to and thereby confirm the (politically controversial, in the United States) link between greenhouse gases and global temperature increase.

The first problem this inference faces is the low probability that extant GCMs constitute a "sufficiently heterogeneous" set, even given the phrase's intensional flexibility. Given the complexity of global climate dynamics and the models developed (very recently) to represent them, there is every reason to suspect that the vast space of representational possibility has been only meagerly sampled thus far (Parker 2011). In fact, many GCMs have not been developed independently and instead descend from a few early models (Parker 2006). This ancestry may explain the systematic errors many GCMs share. As such, the crucial concern that without sufficient model diversity the discovery of a robust property might "depend in an arbitrary way on the set of models analyzed" (Weisberg 2006, 737) has not been alleviated for GCMs. Nor is this worry unique to GCMs. It besets all complex models, the staple currency of theorizing in contemporary ecology, environmental science, and most quantitative sciences.

The second problem is the intractability of GCMs. Determining whether a C exists, formulating and verifying robust theorems, and evaluating their stability properties requires that the model's structure and dynamics be scrutable. It must be clear what components of climatic processes the model captures, which ones are driving dynamics and to what degree, whether properties and relationships are fragile or resistant to disturbing forces, and

so on. The simple predator-prey models Weisberg considered were transparent in this way: the first-order differential equations could simply be inspected and algebraically manipulated to ascertain the relevant robustness desiderata. GCMs are very different. The complexity and sheer number of partial differential equations they involve precludes such inspection and standard analytic solution techniques. As Parker (2009, 233) concisely puts it, “Today’s state-of-the-art climate models are among the largest, most complicated computer simulation models ever constructed.” For this reason, GCMs are solved by computational simulation. But computational solutions provide less insight into model properties than analytic solutions. For example, solutions by simulation usually do not yield a complete survey of all possible solutions. Given that some solutions are unknown, determining the putative C (and thus any case 2 inference) is therefore problematic. For such complex models, the term ‘solved’ is also a bit misleading. Apart from being analytically intractable, these models are usually also computationally intractable: they cannot be directly solved given the computational resources and temporal constraints available. The computational methods themselves simplify the models before solving them and often employ heuristic shortcuts to make the computations manageable (see Winsberg 2001). With each step away from simple, analytically tractable models, the prospect of achieving the four components of Weisberg’s account of robustness analysis decreases.⁴

5. Bayesian Robustness Analysis. Beyond facilitating transmission of empirical support between robust properties and model cores, robustness itself is sometimes considered confirmatory: “Weisberg is appealing to a variety of evidence argument here, because he is appealing to a range of instances of fit of the model over different parameter values, parameter spaces or laws. It is against this background of differing model constructions that the common structure occurs and causes the robust property to appear, and it is the degree of this variety of fit for which the model has been verified that determines how confident we should be in the causal connection” (Lloyd 2009, 221; 2010, 981). Specifically, robustness analysis (case 2: $R \rightarrow C$) is taken to

4. These two problems pose serious obstacles to applying robustness analysis to a relatively new, promising, and computationally intensive approach to modeling ecosystems: individual-based modeling (see Grimm and Railsback 2005). A typical individual-based model is a complex computer-based mathematical model composed of a set of variables for each organism in the ecosystem being modeled and sets of equations representing how each organism consumes resources, develops, interacts with its environment and neighbors, reproduces, and, when applicable, rears its offspring. Individual-based models are incredibly complicated, often involving thousands of variables, parameters, and equations of a variety of mathematical forms. Their dynamics are therefore not analytically tractable and must be investigated by simulation techniques.

be a form of variety-of-evidence inference and “since a variety of evidence does ordinarily give us reason to increase the degree of confirmation of a model, it does in this case as well” (Lloyd 2010, 982).

The “range of instances of fit” refers to the diverse predictive successes of GCMs. In general, each GCM relatively reliably predicts other climatic variables and patterns apart from global mean temperature. According to Lloyd (2009), examples include patterns of precipitation, wind, ocean temperatures, ocean currents, rain belts, monsoon seasons, and troposphere height. This diversity of predictive success is taken to increase confirmation of the “causal connection,” the causal relationship between core greenhouse gas and global temperature the GCMs share.

Whether these predictive successes confirm the relevant models cannot be assessed here (see Parker 2009). Our focus is whether a variety-of-evidence argument—the Bayesian account due to Fitelson (2001) that Lloyd invokes in particular—can show robustness is itself confirmatory. The account relies on the notion of confirmational independence (CI).

(CI) E_1 and E_2 are *confirmationally independent* regarding hypothesis H (with respect to confirmation function c) if and only if $c(H, E_1|E_2) = c(H, E_1)$, and $c(H, E_2|E_1) = c(H, E_2)$ (Fitelson 2001).

Variables E_1 and E_2 designate different bits of evidence; H designates a hypothesis, which Lloyd (2009) takes each GCM to be about global climate; and $c(H, E_i)$ designates the degree E_i confirms H . Note that CI does not require that E_1 and E_2 themselves be logically or probabilistically independent. That many of the targets of predictive success mentioned above are not independent (e.g., precipitation and ocean temperature) is therefore consistent with CI.

With this notion, the confirmational significance of evidential diversity (CSED) can be stated.

(CSED) If E_1 and E_2 individually confirm H , and if E_1 and E_2 are CI regarding H , then $c(H|E_1 \& E_2) > c(H|E_1)$, and $c(H|E_1 \& E_2) > c(H|E_2)$ (Fitelson 2001).

Attempting to apply the account reveals immediate difficulties. First, precisely because relationships exist between the factors GCMs make predictions about, it is unclear whether CI holds. For example, one would expect that models that correctly predict patterns in ocean temperatures would more likely correctly predict patterns in ocean currents than would models that did not. These dependencies, in turn, would almost certainly affect confirmation relationships such that CI would be violated. Specifically, in such cases it seems $c(H, E_1|E_2) < c(H, E_1)$. Most extant statistical characterizations of the

confirmatory value of diverse evidence involve similar or stronger independence conditions (e.g., Franklin and Howson 1984; Earman 1992), so this poses a formidable challenge for this approach to establishing that robustness is confirmatory.

Second, CSED's focus is an individual model (i.e., H), not parts thereof (i.e., the core common structure C). But with this focus, robustness as an inferential tool grounded in properties of models (plural) plays no role. Perhaps the idea is that CSED can be generalized to the relationship between the GCM_i and C . Because each GCM is confirmed by various predictions, perhaps they can be treated as bits of evidence for the common core C that they share. Returning to Fitelson's account and making the relevant **substitutions**, the generalization would require what follows:

If GCM_1 and GCM_2 individually confirm C and are CI regarding the (core) hypothesis C , then $c(C|GCM_1 \& GCM_2) = c(C, GCM_1)$, and $c(C|GCM_2 \& GCM_1) = c(C, GCM_2)$.

But this is flawed on many fronts. First, since C is part of GCM_i , the right side of each equation seems to be 0, and the first part of the preceding antecedent, false: GCM_i deductively entails C , but that certainly does not establish that it confirms C . And, second, since GCM_i and GCM_j ($i \neq j$) are logically incompatible hypotheses about global climate, the left-hand side of each equation seems undefined: the conditionalizations are predicated on an impossible circumstance. The discordance is caused by the shift from a scenario in which a variety of evidence confirms a single model (e.g., a particular GCM) to a multimodel context in which the aim is confirming a model part (e.g., C) via properties of many models (e.g., extant GCMs). The cogency of Fitelson's (2001) account of CSED for the former does not redound to the latter.⁵

6. Conclusion. This analysis accords with the recent largely negative assessment in the literature that robustness is not itself confirmatory (e.g., Woodward 2006; Odenbaugh and Alexandrova 2011; Parker 2011). For example, Parker (2011) considers a different, non-variety-of-evidence Bayesian account of robustness in which R being jointly derived from several models (e.g., GCMs) is itself evidence for, and thus confirms, R . Let e designate this fact of predictive agreement, and let R designate, for instance, the shared prediction that mean global temperature will be 1°–2°C warmer in the 2090s than

5. Note that Lloyd's (2010, 977) description of Fitelson's account correctly focuses on a (singular) model being confirmed, while Weisberg's account of robustness analysis is correctly described in the context of (plural) models. The difficulty of integrating the two is mirrored in the tension between 'the model' and 'differing model constructions' in the passage quoted from Lloyd (2010, 981) at this section's beginning.

in the 1890s. Parker argues the crucial issue is whether $p(e|R) > p(e|\neg R)$, or $p(e|R) \gg p(e|\neg R)$ for significant confirmation, which she argues GCMs are not yet plausible enough to establish.

But if this Bayesian account is satisfactory, the general prognosis is not irredeemably negative.⁶ On this account, robustness could be confirmatory even if GCMs do not yet justify the inference. Moreover, that there are defensible forms of variety-of-evidence arguments codified within well-developed statistical frameworks, and that inferential robustness seems to use diverse models in an evidentially similar way, suggests that a cogent but more complicated statistical basis for the confirmational value of robustness will eventually be found. The models on which robustness is predicated will likely never be logically or statistically independent, so the fact that at least one account, Fitelson (2001), does not require such independence, only confirmational independence, is cause for cautious optimism. Statistical bases for legitimate scientific methods often emerge only after the latter are common practice.

REFERENCES

- Earman, John. 1992. *Bayes or Bust?* Cambridge, MA: MIT Press.
- Fitelson, Branden. 2001. "A Bayesian Account of Independent Evidence with Applications." *Philosophy of Science* 68:123–40.
- Forber, Patrick. 2012. "Modeling Scientific Evidence: The Challenge of Specifying Likelihoods." In *EPSA Philosophy of Science: Amsterdam 2009*, 55–66. Dordrecht: Springer.
- Franklin, Allan, and Colin Howson. 1984. "Why Do Scientist Prefer to Vary Their Experiments?" *Studies in History and Philosophy of Science* 15:51–62.
- Grimm, Volker, and Steven Railsback. 2005. *Individual-Based Modeling and Ecology*. Princeton, NJ: Princeton University Press.
- Levins, Richard. 1966. "The Strategy of Model Building in Population Biology." *American Scientist* 54:421–31.
- . 1993. "A Response to Orzack and Sober: Formal Analysis and the Fluidity of Science." *Quarterly Review of Biology* 68:547–55.
- Lloyd, Elizabeth. 2009. "Varieties of Support and Confirmation of Climate Models." *Proceedings of the Aristotelian Society* 83:213–32.
- . 2010. "Confirmation and Robustness of Climate Models." *Philosophy of Science* 77:971–84.

6. Concerning this account's tenability, a nontrivial worry is worth noting: $p(e|R) > p(e|\neg R)$ is not an inequality between garden-variety likelihoods. Typical likelihoods in science concern specific hypotheses that models are constructed to express. A model typically generates a probability distribution over possible observations that determines the relevant likelihood. But in this case, the "hypothesis" R is a prediction about future mean global temperature, and the "evidence" is a congruence between models. With these non-standard specifications, there seem to be no resources for estimating probabilities, particularly for this congruence; the catchall $\neg R$ is particularly uninformative in this regard. Rather than capture how robustness should be represented within a Bayesian framework, the worry is that ill-defined conditional probabilities have been assumed to do work that they cannot. This broaches difficult issues about how likelihoods are determined in science that cannot be addressed here (see Forber 2012).

- Odenbaugh, Jay, and Anna Alexandrova. 2011. "Buyer Beware: Robustness Analysis in Economics and Biology." *Biology and Philosophy* 26:757–71.
- Orzack, Steven, and Elliott Sober. 1993. "A Critical Assessment of Levins's *The Strategy of Model Building in Population Biology*." *Quarterly Review of Biology* 68:533–46.
- Parker, Wendy. 2006. "Understanding Pluralism in Climate Modeling." *Foundations of Science* 11:349–68.
- . 2009. "Confirmation and Adequacy-for-Purpose in Climate Modeling." *Proceedings of the Aristotelian Society* 83:233–49.
- . 2010. "Whose Probabilities? Predicting Climate Change with Ensembles of Models." *Philosophy of Science* 77:985–97.
- . 2011. "When Climate Models Agree: The Significance of Robust Model Predictions." *Philosophy of Science* 78:579–600.
- Sober, Elliott. 1993. "Mathematics and Indispensability." *Philosophical Review* 102:35–57.
- Trout, J. D. 1993. "Robustness and Integrative Survival in Significance Testing: The World's Contribution to Rationality." *British Journal for Philosophy of Science* 44:1–15.
- Weisberg, Michael. 2006. "Robustness Analysis." *Philosophy of Science* 73:730–42.
- Wimsatt, William. 1981/2007. "Robustness, Reliability, and Overdetermination." In *Re-engineering Philosophy for Limited Beings*, 37–74. Cambridge, MA: Harvard University Press.
- Winsberg, Eric. 2001. "Simulations, Models, and Theories: Complex Physical Systems and Their Representations." *Philosophy of Science* 68 (Proceedings): S442–S454.
- Woodward, James. 2006. "Some Varieties of Robustness." *Journal of Economic Methodology* 13: 219–40.