Robust! Handle with Care

Wybo Houkes and Krist Vaesen*†

Michael Weisberg has argued that robustness analysis is useful in evaluating both scientific models and their implications and that robustness analysis comes in three types that share their form and aim. We argue for three cautionary claims regarding Weisberg's reconstruction: (1) robustness analysis may be of limited or no value in evaluating models and their implications; (2) the unificatory reconstruction conceals that the three types of robustness differ in form and role; (3) there is no confluence of types of robustness. We illustrate our central first claim with a case study: the application of Lotka-Volterra models to technology diffusion.

1. Introduction. Philosophical work on robustness analysis has its starting point in works by Richard Levins (1966) and William Wimsatt (1981). Both authors elaborate in different ways the claim that, although applications in the sciences may differ, evaluating modeling efforts in terms of 'robustness' has the core meaning of identifying which properties or results are shared by various models or various specifications of the same model. In Levins's oft-quoted words, robustness analysis is a method to find "truth at the intersection of independent lies." Moreover, both authors suggest that robustness provides warrant for believing a result or a nonobservational source of confirmation of the result.

Later authors have argued against the unity of robustness analysis and its confirmatory power. Jim Woodward (2006), taking this critical approach, urges that we need to distinguish several varieties of robustness, which differ in their target. Inferential robustness, for instance, concerns conclusions

Received March 2011; revised November 2011.

*To contact the authors, please write to: Philosophy and Ethics, School of Innovation Sciences, Eindhoven University of Technology, Den Dolech 2, NL-5600 MB, Eindhoven, The Netherlands; e-mail: w.n.houkes@tue.nl; k.vaesen@tue.nl.

†Research by Wybo Houkes and by Krist Vaesen was made possible by the Netherlands Organization for Scientific Research (NWO). An earlier version of this article was presented at ECAP7, Milan, September 2011. We would like to thank, with the usual disclaimer, two anonymous referees for their helpful comments.

Philosophy of Science, 79 (July 2012) pp. 345–364. 0031-8248/2012/7903-0002$10.00
Copyright 2012 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/666061 Published online by Cambridge University Press

H drawn from (typically, hypotheses entailed by) data sets, in combination with various assumptions incorporated in a model. Derivational robustness, by contrast, concerns the derivation of observable results from models, to test how sensitive the model's capacity for yielding accurate predictions is to, for instance, parameter settings. This differentiates robustness analysis as a way to evaluate model implications from evaluating models themselves as robust. Woodward combines this differentiation with doubt regarding any evaluative use of robustness analysis. Following Orzack and Sober (1993), Woodward gives the formal argument that if and only if a complete set of assumptions entails the same hypothesis, inferential robustness warrants belief in a hypothesis. Since completeness is practically unattainable, and it is unclear how standards could be lowered without losing warrant for belief, Woodward concludes that inferential robustness has no confirmatory power and little scientific use, apart from revealing hidden and possibly testable assumptions shared by otherwise diverse models—a role that is also discerned by Forber (2010).

A defense of robustness analysis as a unified and evaluatively useful procedure has been given by Michael Weisberg. In two recent papers, Weisberg (2006; Weisberg and Reisman 2008) argues that robustness analysis helps to identify *robust theorems*: conditional claims that hold independently from simplifying assumptions and other structural features of the model(s). In robustness analysis, one tests a group of models of the same phenomenon: if all yield the same result, one can identify the common structure in these models and conclude—after two more steps in a four-step procedure—that "ceteris paribus, if [common structure] obtains, then [robust property] will obtain" (Weisberg 2006, 738).

Weisberg and Reisman (2008) distinguish three types of robustness, all of which are aimed at identifying robust theorems of this conditional form. In *parameter* robustness, the theorem holds over different parameter settings; in *structural* robustness, the theorem holds when changes are made to the causal structure of the system being modeled; and in *representational* robustness, the theorem holds irrespective of the representational framework used.

Robust theorems and the procedures identifying them are scientifically interesting because they allow "determining which models make trustworthy predictions and which models can reliably be used in explanations" (Weisberg 2006, 731). This evaluative role of robustness analysis is especially useful in the absence of other ways to determine the distortive effects of modeling assumptions and idealizations. In particular, if a set of models with a common structure is analyzed and found to have robust implications, it is "very likely that the real-world phenomenon has a corresponding causal structure" (739). Conversely, if there is "low-level confirmation" (sec. 4) of the models in question, "robust theorems make claims about real-world

https://doi.org/10.1086/666061 Published online by Cambridge University Press

phenomena." Thus, Weisberg captures robustness analysis in a four-step procedure that ends in theorems with a common form, and he saves the evaluative role of robustness regarding models (in case the existence of a common causal structure is concluded) and their implications (in case the existence of a real-world phenomenon is concluded).

This unificatory reconstruction of robustness analysis adds considerable clarity to the original views of Levins and Wimsatt.[1] Moreover, it appears to address abstract criticisms like Woodward's, by focusing on actual applications of robustness analysis, illustrated with the role of robust theorems in population ecology. However, the reconstruction is not unproblematic. Considerable caution is required in applying robustness analysis in the unified form and evaluative role of Weisberg's reconstruction. We shall argue for three cautionary—occasionally critical—claims:

1. Robustness analysis may be of limited or no value in evaluating models and their implications.
2. The unificatory reconstruction conceals that Weisberg's three types of robustness differ in form and role.
3. There is no confluence of types of robustness: their results cannot be combined into a relevant robust theorem.

Weisberg's reconstruction is not incompatible with these claims; hence, we present them primarily as cautionary remarks or words of warning against misinterpretations or misapplications. Yet Weisberg explicitly denies claim 2 with regard to the role of representational robustness; some recent applications of his analysis by others (e.g., Lloyd 2010) involve denials of claim 1, and the unified form of Weisberg's reconstruction misleadingly suggests confluence.

In section 2, we establish claim 1, which is central to our arguments. First, we use Weisberg's four-step reconstruction of robustness analysis to identify two methodological choices: that of a model family, of which the behavior is studied, and that of converting the descriptive (behavior-comparative) result of robustness analysis into an evaluative claim about either models or their implications. We then present arguments, separately for the evaluation of models and of their implications, to the effect that a crucial role is played by the required choice of a credible model family and that this role limits and potentially minimizes the evaluative impact of robustness analysis. The resulting tension between the choice of a credible model family and an evaluative role of robustness analysis replaces the trade-off between completeness and robustness that is noted by more formalist critics.

1. We refer to Weisberg's account as a "reconstruction" and reserve "analysis" for robustness analysis.

https://doi.org/10.1086/666061 Published online by Cambridge University Press

Our first cautionary claim is illustrated with a case study in section 3. This involves the same modeling approach—Lotka-Volterra models—that Weisberg uses as an illustration, applied not to ecological predator-prey systems but to the diffusion of technological innovations. We first show how, in this domain, some investigations regarding model behavior fit Weisberg's reconstruction of robustness analysis. Then, we argue that the evaluative impact of these investigations is strongly qualified by the prior credentials of Lotka-Volterra models in this area, in line with our main cautionary claim.

In section 4, we derive the other two claims as corollaries of the first. We argue that parameter and structural robustness are distinguished by different prior choices of model family, which blocks confluence of their results. Moreover, we argue that the results of parameter robustness can in general only be trivially cast in Weisberg's form of a robust theorem, because no previously unknown common structure is discovered, and that representational robustness is apparently not characterized by prior choice of a model family and can only play a model-constructive or explorative role.

2. Family Matters. Weisberg (2006, sec. 3.2) reconstructs robustness analysis as a procedure with four conceptually distinct, but not necessarily consecutive steps:

i) finding *robust properties*—behavior shared by all members of a group of models
ii) analyzing which structural features of the models generate the shared behavior—leading to formulation of a *robust theorem*
iii) *empirical interpretation* of the robust theorem
iv) analyzing the stability of the robust theorem to find *defeating conditions*.

Here, step i involves finding a set of implications of the form

$$(M_1 \,\&\, A) \rightarrow H$$
$$(M_2 \,\&\, A) \rightarrow H$$
$$\ldots,$$

where M_1 and M_2 are models, H is a hypothesis entailed, and A is a set of data and auxiliary assumptions needed to derive the shared implication.[2] Consequently, the result of step i, identification of a robust property H, is an implication of the form

2. Variable A is introduced to prevent suggesting that models entail hypotheses in and of themselves. We assume that a shared set of auxiliary assumptions and data can be specified for each robust property entailed—a strong assumption, but it plays no role in our analysis.

https://doi.org/10.1086/666061 Published online by Cambridge University Press

$$(\text{RP}^*)(\{M_1 \vee M_2 \vee \ldots\} \& A) \rightarrow H.$$

Sometimes, this list contains a countable number of discrete items, for instance, in investigating whether nonlinear oscillators persist in showing oscillatory behavior when a disturbance term is introduced. In other cases, a range of models is investigated, for instance, in seeking the maximum value of a disturbance term under which oscillators show oscillatory behavior. Conceivably, the list could be open-ended, for instance, when studying the influence of disturbance terms in general.

Weisberg (2007) follows van Fraassen in characterizing a model as a mathematical structure—fundamentally, a collection of trajectories in state space. In line with this, steps i and ii of the procedure are mathematical operations and lead to a theorem of the form

> (RT*) For all models on a considered list, if a model has the mathematical structure *X**, then it entails hypothesis *H*.

In other words, robustness analysis concerns properties of models, or hypotheses (statements) entailed by models, rather than connections between causal structures and real-world phenomena. The third step, that of "empirical interpretation," converts RT* into

> (RT) Ceteris paribus, if common causal structure *X* obtains, then robust property *H* will obtain.

Here, "common" refers to the considered list of models and interprets the mathematical structure *X** shared by some models as a causal structure *X*. In principle, the earlier steps may also reveal that several models on the list do not entail *H* and do not have *X**. This means that robust theorems, mathematical as well as causal, are descriptive and implicitly comparative: robustness analysis identifies a connection between the structure and implications of some models and might simultaneously identify models for which this connection does not hold.

Consequently, two questions arise with regard to the evaluative role of robustness analysis. The first is, which models are chosen as candidates for the list? The second is, how can a description and comparison of the behavior of models play a role in evaluating models and their implications?

Formalist critics of robustness analysis have clear answers to both questions: all conceivable models (of a phenomenon) should be considered as candidates for the list, and if and only if this is done and some shared implication *H* is found, robustness analysis provides warrant for belief in *H*. Evidently, it is difficult to consider all conceivable models and unlikely that

https://doi.org/10.1086/666061 Published online by Cambridge University Press

there are any similarities to be found on such a complete list. The resulting robustness-completeness trade-off (Woodward 2006) dispels any realistic hope of finding an evaluative role for robustness analysis.

Weisberg dismisses this line of argumentation as too abstract and seeks an evaluative role for robustness, even in cases in which an incomplete list of models is considered. This means that he must provide alternative answers to both questions given above. The four-step reconstruction suggests where answers are required, namely, in step iii, in moving from a descriptive to an evaluative result of robustness analysis, and before step i, in choosing what we shall call a *model family*.

The answers given by Weisberg are not as clear as his reconstruction of the procedure of robustness analysis. He suggests that, rather than a complete list, one ought to consider "a group of similar, but distinct models" or a "sufficiently diverse" or "sufficiently heterogeneous" set (2006, 737, 739) and that the transition to evaluation is supported by "low-level confirmation."[3] These answers are minimally in need of further clarification. Some is provided implicitly by Weisberg's illustration of robustness analysis in population ecology. We take an explicit and systematic tack here. Looking at the evaluations of models and their implications in turn, we reveal how the evaluative power of robustness analysis is limited by the required choice of a credible model family, that is, by the first of the two issues identified above. While this does not make it impossible to salvage the belief that "robustness . . . is a Good Thing" (Woodward 2006, 219), it reveals that robustness analysis by itself does not determine how "Good" a robust model or robust theorem is.

2.1. Evaluating Models. We start on the 'model-side' of the robust theorem and consider confirmation of instantiation of X—which might reflect positively on the explanatory value of models $\{M_1, M_2, \ldots\}$. Suppose that one actually observes H in some target system. Then, how likely is it that the causal structure X is instantiated in that system? Without going into confirmation-theoretical details—which we are not equipped to give—it appears that the credibility of X is determined not only by the diversity of models $\{M_1, M_2, \ldots\}$ analyzed but also by the choice of model family. The diversity of family members may be constrained before any analysis of their individual behavior, up to the point where the shared structure is a result of choice of family rather than an outcome of the analysis. So at which stage of the analysis and on what grounds is the model family chosen?

3. "Low-level confirmation is what allows robust theorems to make claims about real-world phenomena" (Weisberg 2006, 740).

https://doi.org/10.1086/666061 Published online by Cambridge University Press

One strategy is to choose the family before step i. This makes any restrictions on the family into an input of robustness analysis; that is, only models within those restrictions are considered throughout the four steps. Then, instantiation of X is only a credible result of robustness analysis provided H is observed and the restrictions are credible independently of robustness analysis. This requires warrant for believing that an adequate model may be found within a family, without having grounds for preferring one of the family members over another. If there were such grounds, robustness analysis would have no evaluative power: we would already know which model (in the family) is the most adequate. If family members are equally credible, robustness analysis could add to the family credentials an identification of shared behavior and, perhaps, a shared structure responsible for this behavior. What it does not—cannot—add is credibility for the model family and, therefore, warrant for the instantiation of the causal structure. Roughly put, the causal structure is only as robust, and the explanatory power of considered models only as large, as the model family considered is credible.[4] Furthermore, the result of robustness analysis would be more accurately formulated as

> (RT′) For all systems represented by a model family, ceteris paribus, if common causal structure X obtains, then robust property H will obtain.[5]

which limits the possibility to evaluate models by means of robustness analysis.

As a first brief illustration of these limits, consider model systems of coupled harmonic oscillators moving on a frictionless plane. These systems may comprise a credible model family, and one could study which behavior is shared by models with different values of the force constants of the connecting springs. Then, robustness analysis may reveal shared behavior of models with some values for force constants and even some underlying mathematical structure. Belief in the causal counterpart of this structure, in case the shared behavior is actually observed, is only warranted insofar as

4. In statistics, credentials of model families are typically established by formal criteria such as the Bayesian information criterion or Akaike's information criterion (see, e.g., Forster and Sober 1994; Kieseppä 2001). For instance, one model family could be the class of all linear models, whereas another could be the class of all logarithmic models. The Bayesian and Akaike's information criteria would then help to identify the family (linear or logarithmic) that accords best with the available observations. We believe that the statistical notion is useful for demarcating model families but that their credentials must be construed broader, at least in the context of the present article. Credentials refer to the match between model and data and to the credibility of the causal mechanisms that need to be assumed to get this match (for an illustration of this, see n. 16).

5. We thank an anonymous referee for suggesting this formulation.

https://doi.org/10.1086/666061 Published online by Cambridge University Press

the frictionless-plane family was credible in the first place. Moreover, one may be warranted in preferring those models over other family members but not in considering these models validated tout court. Robustness analysis cannot confirm its own constraints: it makes the model family no more credible than it was before the analysis. Even in the unlikely case that all frictionless coupled-oscillator models would share some behavior that is actually observed, the family does not gain credibility.

An alternative to this prior-choice strategy is to identify the model family as a result of robustness analysis. Here, robustness analysis is used to explore model space without constraints, with an eye to identifying similarities and differences in model behavior. This might reveal a variety of models compatible with an empirical result. This strategy has the advantage that any common structure identified in these models does not depend for its credibility on the prior credentials of some model family. Yet the lack of restriction in that respect also means that instantiation of the causal structure is only credible if, as the formalist critics submit, all conceivable models of a target system or phenomenon are inspected. Until this completeness has been achieved, any shared structure found responsible for a robust property may be an artifact of the limited scope of explorative robustness analysis, even if the implications of highly diverse models would be inspected. Therefore, this explorative strategy provides no warrant for accepting models: their explanatory value, in identifying a causal structure that may be responsible for the robust property observed, may likewise be an artifact of the scope of exploration. Another reason why explorative robustness analysis has no evaluative power regarding models is that it may yield that the inspected property is entailed by models that do not share any structure. Then, robustness analysis ends at step i and reveals underdetermination of adequate models by the target phenomenon. This is a scientifically interesting result, and it is satisfactory if the goal was to explore but not if it was to evaluate the explanatory value of models: the question posed above—"How likely is it that causal structure *X* is instantiated in a target system, given observation of *H*?"—cannot be answered in this case, for lack of an *X* to instantiate.

It is worthwhile to consider, against this background, Weisberg's notion of "low-level confirmation" (esp. Weisberg 2006, sec. 4), for it indicates how prior family credentials might curtail the scope of robustness analysis or guide the interpretation of RT*.[6] According to Weisberg, low-level confirmation concerns "the fact that certain mathematical structures can adequately represent properties of target phenomena" (740). This means that it might reflect family credentials as well as support the empirical

6. We consider these and alternative readings of the notion primarily to strengthen our first cautionary claim, not to find the most charitable reading of Weisberg's remarks on low-level confirmation.

https://doi.org/10.1086/666061 Published online by Cambridge University Press

interpretation of shared model implications: it entails that certain structures and not others are representative.

Now there are several options with regard to the scope of low-level confirmation. The first is that it applies to a broad mathematical framework, say that of coupled differential equations. This leaves significant scope for robustness analysis. However, it also entails that robustness analysis is merely explorative, as long as one only considers the behavior of specific systems of coupled differential equations (say, Lotka-Volterra equations): additional credentials must be provided to warrant this choice before it supports evaluative claims. Moreover, it is not clear how the general credentials of a mathematical framework would facilitate the empirical interpretation of RT*—which is the role that Weisberg envisages for low-level confirmation. A second option is therefore to have low-level confirmation apply to a specific model family, such as Lotka-Volterra equations. Then, robustness analysis is much more practicable, and empirical interpretation of RT* mainly concerns application to particular predator-prey systems, say populations of sharks and cod. Yet in this case, the evaluative potential of robustness analysis is circumscribed since it assumes and cannot improve the credentials of Lotka-Volterra models. Finally, low-level confirmation might apply to individual models. In that case, it determines empirical interpretation and warrants strong limitations on the models that enter on the list in RP*. However, it deprives robustness analysis of any evaluative power regarding these models.[7]

To sum up, robustness analysis requires a prior choice of model family to play any evaluative role regarding models. This role is then curtailed by the prior credentials of the model family. Alternatively, robustness analysis may be used to explore similarities in model space. On either option, the evaluative power of robustness analysis is limited and possibly minimal.[8]

An easier route to the same conclusions is provided by a practice-based view of modeling. There, it is emphasized (as in Weisberg 2007) that the behavior of models is often studied apart from considering whether this

7. Weisberg (2006, 740–41) suggests that low-level confirmation is based on predictive accuracy and that it warrants belief in the representational accuracy of both "the mathematics of the logistic model" and "the models described by coupled differential equations." This illustrates, in our opinion, the ambiguities of scope in the notion of low-level confirmation: even if one assumes that the former applies to all models described by the logistic equation, it is much more specific than the latter—and one would expect such differences to be relevant to the scope of robustness analysis.

8. Applications of robustness analysis to evaluate models may be a hybrid of the two strategies mentioned: they start with some rough idea of a credible model family and use robustness analysis to focus on both the family and the 'robust' models within the family. This hybrid strategy can be reconstructed as an evaluative application of robustness analysis with prior choice of the model family that was actually discovered by exploration.

https://doi.org/10.1086/666061 Published online by Cambridge University Press

behavior represents anything in the target system. However, scientists only study the behavior of a model for which they have some reason—no matter how weak and defeasible—to expect a measure of predictive or explanatory adequacy. This expectation is typically part of the construction of a model.[9] This, in turn, leads to two possibilities for choosing the model family $\{M_1, M_2, \ldots\}$ that, if a shared implication is found, features in RP*. One is that the behavior of various models is studied, for which one has independent adequacy expectations. Alternatively, one might study only one model, identify an interesting implication, and vary parameter settings and structural features—deliberately constructing a set $\{M_1, M_2, \ldots\}$. In this case, construction of alternative models should not involve defeaters of the adequacy expectation: although one might consider the implications of less credible models than M_1 as an academic exercise (e.g., modeling predator-prey systems with the heat equation), finding that these models do not entail H is not likely to lead one to conclude that H is not robust. Hence, prior expectations regarding the adequacy of models determine the scope of robustness analysis and cannot themselves be confirmed.

2.2. Evaluating Results. We now turn to the 'results-side' of robustness analysis. Can robustness analysis identify empirical regularities? That is, how likely is it that, given auxiliary assumptions, if the causal structure X is instantiated in real-world systems, the robust property H obtains?

Here, the diversity of the members of the model family may be advantageous since a hypothesis is more strongly confirmed by instances of different sources of evidence than by an equal number of instances from the same source. One might intuit that this touchstone of confirmation theories also applies to confirmation of observable hypotheses by similar and different models. Here, similarity and difference refer to the *evidential basis* (possible or actual) of models. So, again, the choice of model family may have a significant impact on the significance of robustness analysis. Consider again, the form of the robust theorem presented above:

> (RT′) For all systems represented by the model family, ceteris paribus, if common causal structure X obtains, then robust property H will obtain.

This conditional form brings out how robustness analysis might be a procedure for identifying highly circumscribed empirical regularities. Confirmation of such regularities is determined by two factors minimally: the credentials of the family of models considered, which must now be an input of robustness analysis, and the diversity of the models within the family, which is the result of robustness analysis. The relative weight of these

9. Boumans (1999) refers to this as the "built-in justification" of models.

https://doi.org/10.1086/666061 Published online by Cambridge University Press

factors cannot be determined without considering the details of specific applications, but it would seem that the first factor may outweigh the second: if the models considered were not too credible to begin with, demonstrating that they, despite their diversity, have a result in common does not lend much credibility to the result. Furthermore, the weight of the second factor increases, the more models are considered, up to the point where there are no prior restrictions on the model family. In the limit, the situation reduces to that considered by the formalist critics: the result is known to be true, once all possible models are considered; until that, that a result is shared may be due to the limited scope of the analysis or to hidden but questionable assumptions shared by the models.[10]

What can be stated in general is that robustness analysis can only be used to evaluate model implications relative to a choice of model family. Then, one should have reason to think that models in this family, no matter how diverse they might otherwise be, are credible before their similarities become interesting. Within these restrictions, robustness analysis might provide some additional reason to believe predictions that are shared by diverse models, meaning that the confirmatory role of robustness analysis is nonmarginal in specific cases. However, if the remaining diversity within a model family is small given the restrictions, the robust result adds very little to the family credentials. Thus, in general, robustness may only marginally warrant belief in an implication over and above the low-level confirmation of the family. Any specific case, such as the implications of climate models discussed by Lloyd (2010), could exemplify either the marginal or the nonmarginal scenario, depending on the confirmation of the model family and the diversity of family members. Only a detailed reconstruction, including an operationalization of diversity, could reveal whether robustness of an implication adds anything to its credibility. We do not presume that these details are impossible to specify or that results can only be disappointing for advocates of robustness analysis. Our discussion does show, however, that merely pointing out that different models share an implication may be evaluatively insignificant—providing grounds for caution in each specific case.

10. Weisberg (2006, 742) writes that low-level confirmation, not robustness analysis, provides confirmation of robust theorems. This might underestimate the evaluative significance of robustness analysis: if low-level confirmation, as suggested above, applies to the model family, robustness analysis may provide additional confirmation for an implication shared by all members of the family. In Weisberg's example, robustness analysis would warrant belief in the empirical regularity represented by the Volterra principle—but only for systems that one believes (on independent grounds) are represented by Lotka-Volterra models. It cannot, as argued above, provide additional warrant for the latter belief—but low-level confirmation might, at least in the second of the three interpretations considered earlier.

https://doi.org/10.1086/666061 Published online by Cambridge University Press

3. A Case Study: Lotka-Volterra Models of Technology Diffusion. As an extended illustration of our main cautionary claim, we look at a particular modeling practice: the application of Lotka-Volterra models to processes of technology diffusion. This case involves models that are mathematically similar to those used as an illustration by Weisberg. Yet their application in an unfamiliar (for philosophers) and not immediately receptive (for researchers of technology diffusion) context brings to light the importance of family credentials in robustness analysis. We start by briefly describing the disciplinary context for this application of Lotka-Volterra models and then illustrate our claim.

Perhaps surprisingly, given the complexity of the phenomenon, there are several models that accurately predict the diffusion rates and, consequently, product sales of new technologies. The earliest and still most successful work started from the observation that, as soon as innovations have acquired a small market share, their growth rates follow a sigmoid (S-shaped) curve.[11] One model of this process fits a linear form of the logistic curve to the growth rates—for 17 cases in the original paper (Fisher and Pry 1971) and around 100 since then. Alternatives offered in the literature include the Bass (1969) model, which divides a population of consumers into innovators and imitators with different propensities to adopt an emerging technology, and various Gompertz models, in which a new technology replaces a mature technology that gradually falls into disrepair. The predictive success of these models has led to widespread adoption in industry, and research in technology forecasting has increasingly focused on hybridizations of existing models rather than on proposing alternatives.[12]

Gradual refinement of the phenomenological models used for forecasting has been accompanied by attempts to construct more explanatory models. One such attempt follows a suggestion in Fisher and Pry (1971) and seeks to understand the diffusion of technologies as, primarily, a process of competition between an emerging and an established technology. Several researchers have, for explicitly explanatory purposes, sought to apply the Lotka-Volterra Competition (LVC) equations to the growth rates of rival technologies. They describe the merits of these models as providing "clearly defined assumptions about the nature of technological growth" (Porter et al. 1991, 197) and "allowing intuitive understanding of the factors driving substitution" (Morris and Pratt 2003). A general procedure for fitting LVC models to data sets, inspired by modeling practices in ecology, is described

11. Reviews of these and other models may be found in, e.g., Porter et al. (1991) and Michelfelder and Morrin (2005).

12. A detailed comparison of 29 phenomenological models—including several variants of the Fisher-Pry and Bass models—is given by Meade and Islam (1998), who show that a combination provides a better fit to data sets than each of the individual models.

https://doi.org/10.1086/666061 Published online by Cambridge University Press

and applied to four data sets by Farrell (1993).[13] More indirectly, the LVC model has been assessed by studying its relation to the various phenomenological models described above. Morris and Pratt (2003) argue that, under a range of conditions, the LVC model reduces to the logistic Fisher-Pry model but that the model cannot achieve arbitrarily close fits to the Bass and Gompertz models. For these models, similar explanatory underpinnings are available: Henrich (2001) has derived the Bass model from assumptions about social-learning mechanisms, whereas the Gompertz models focus on obsolescence mechanisms.

Some contributions to this body of literature feature investigations into the behavior of LVC models that fit Weisberg's reconstruction of robustness analysis, although these investigations have not led to 'textbook results' on a par with the Volterra theorem (which, interestingly, is never mentioned in this corpus). One example is a result obtained by Saviotti and Mani (1995) for systems of interacting technologies. Saviotti and Mani use the Lotka-Volterra multispecies equation for a finite number of competing technologies.[14] Then, through simulations, they show that technologies with high intrinsic growth rates exhibit chaotic dynamics, on high introduction rates of new technologies. This behavior is interesting because populations "exhibiting chaotic motion do not survive for long" (386)—a result said to be familiar from population ecology. Therefore, it is interpreted as showing that, in systems with high intertechnology competition, "technologies are selected out of the system at least as fast as they are introduced" (388).[15] This property is found for a range of growth-rate values and values of the interaction parameter; it is, in other words, parameter robust. Ceteris paribus clauses are mostly implicit, but one that is generally mentioned in the literature is market stability—akin to the environmental stability assumed in ecological applications. Therefore, on Weisberg's reconstruction, the result obtained by Saviotti and Mani would be of the form

> Ceteris paribus, in systems with high intertechnology competition, technologies are selected out of the system at least as fast as they are introduced.

Now, what is the evaluative impact of this example of robustness analysis with respect to LVC models and to the result derived by Saviotti and Mani?

13. The data sets concern substitution of types of food cans (soldered by lead-free), carpets (woven by tufted), pens (fountain by ballpoint), and tire cords (rayon by nylon).

14. They present the equation in its most general form but do not include mutualistic interactions in their application.

15. High intertechnology competition is defined in the article as a range of values of a parameter that depends directly on the intrinsic growth rate of technologies and the interaction parameters.

https://doi.org/10.1086/666061 Published online by Cambridge University Press

We argue for each in turn that—in line with our reflections in section 2—this impact is, at best, limited.

Suppose, first, that we find high extinction rates of some technologies (say, cell phone models). Does this reflect positively on the explanatory value of LVC models; that is, can we conclude that extinction rates are caused by high intertechnology competition? This conclusion would be premature since there are alternative models—for example, the Bass model or, in case this is regarded as purely phenomenological, Henrich's (2001) rendition of the model in terms of social-learning mechanisms—that might explain high extinction rates without appealing to any mode of competition. Saviotti and Mani's investigations do not encompass the implications of these models but assume that LVC models are credible. The proper form of their result is therefore

> For any real-world system that can be represented by LVC models, ceteris paribus, in systems with high intertechnology competition, technologies are selected out of the system at least as fast as they are introduced.

In other words, this example of robustness analysis (like any example) assumes the credentials of a model family—here, LVC models—and cannot establish these credentials. The same would apply if Saviotti and Mani would have found a common structure that gives rise to the mentioned property and would have interpreted it as a causal structure *X*. Let us call this causal structure "c-mode competition." Then, the theorem would have had the more elaborate conditional form:

> Ceteris paribus, if technologies compete in c-mode, in systems with high intertechnology competition, technologies are selected out of the system at least as fast as they are introduced.

Here, concluding from a high extinction rate—or even from a correlation between a high extinction rate and high intertechnology competition values—that technologies compete in c-mode still assumes that competition (of any mode) is the primary mechanism behind technology diffusion; that is, the theorem is of the RT′ form. This is not to say that the robust theorem is an uninteresting artifact of Saviotti and Mani's exclusive focus on LVC models: if such models are credible, the fact that many of them share particular properties is relevant.

As obvious as it may seem that evaluating the explanatory value of LVC models by robustness analysis shows these limitations, it is equivalent to concluding from an instantiation of the Volterra principle in a real-world system that the corresponding causal structure also obtains: this is only as

https://doi.org/10.1086/666061 Published online by Cambridge University Press

credible as Lotka-Volterra predator-prey models of these systems. If the inference to c-mode competition may seem less credible, it is mainly because the credentials of LVC models of technology diffusion may be weaker, and the issue at hand is not family credentials, per se, but the added value of robustness analysis.

With regard to evaluating the result, a similar conclusion follows. Take the hypothetical structural-robustness formulation of the theorem and suppose that we know that technologies interact through c-mode competition. Can we then conclude that there will be high extinction rates in technological systems with high introduction rates? We can if there is warrant for thinking that competition is the only mechanism that drives technology substitution, that is, in case LVC models are credible. If obsolescence and word of mouth are important mechanisms, the conclusion might not hold.[16] Again, the qualification may seem obvious since it is not too credible that competition is the only mechanism at work. Yet it again shows the relevance of model-family credentials in determining the scope and impact of robustness analysis.

4. Two Corollaries. In section 2, we considered how the evaluative power of robustness analysis is affected by the choice of a model family, and we established a first cautionary claim. In this section, we derive two more cautionary claims as corollaries of the first. These concern the unified and formal character of Weisberg's reconstruction. We argue that, contrary to what is suggested by this reconstruction, the three types of robustness analysis distinguished by Weisberg and Reisman (2008) do not and cannot play the same role in modeling practices and that their results cannot be considered in isolation from domains of application of a modeling technique.

First, our previous reflections provide grounds to suspect that the uniform reconstruction of parameter, structural, and representational robustness overestimates the similarities between these types and—in contrast with the apparent uniformity—conceals differences in both form and role. For a start, the types reflect differences in scope of robustness analysis and therefore differences in model family chosen: in parameter robustness, for instance, one considers members of a family that share their mathematical structure, whereas structural robustness compares models with slightly different structures. These choices of model family may be either an input or the result of the analysis, as discussed in section 2. Yet it is typically not found through robustness analysis that only different parameter settings

16. In other words and as suggested in n. 4, the credentials of LVC models refer not just to their match with observed data but also to the credibility of the mechanisms supposedly responsible for this match. In case competition was known to be irrelevant to technological change, the credentials of LVC models would be low, even if they produced a close fit with observations.

https://doi.org/10.1086/666061 Published online by Cambridge University Press

need to be considered, whereas structural features should be kept invariant. Thus, although Weisberg's classification could be taken as a post hoc statement of results, it is more straightforwardly understood as conveying which model family is chosen before analysis. Indeed, one might replace Weisberg's general parameter-structural classification with more specific characterizations (e.g., parabolic-equation robustness), for the purpose of communicating credentials. The need to choose a model family provides a reason to differentiate specific robustness analyses by their roles since the choice of model family constrains or even minimizes any evaluative role that robustness analysis could play. This cuts across the distinction between parameter and structural robustness since both could involve either a broad or a narrow choice of model family and therefore be significantly or marginally evaluative. That this evaluative significance depends on the details of the case is all the more reason to keep in mind that a unified reconstruction of robustness analysis conceals relevant differences between specific analyses (e.g., concerning scope or role).

A further reason is that it seems contrived, or premature, to present the result of parameter robustness analysis as a robust theorem of the form RT or RT′. In an analysis of the parameter robustness type, one considers similarities between models that share their mathematics aside from parameter settings; it is not to be expected that a mathematical structure X^* may be found that is responsible for similarities in model behavior. In general, results would rather be represented in terms of parameter-value intervals:

> (PRT′) For all systems in a model family that share their mathematics, ceteris paribus, for parameter values in the interval $[p_1, p_2]$, robust property H obtains.

Analysis may reveal that models with these parameter values in fact have some hidden mathematical structure that is not shared by models with parameter values outside the interval. There is no reason to take this as the rule rather than an exception or to assume that finding this hidden structure is the general purpose of parameter robustness analysis: scientific modeling abounds with claims of the form PRT′ and variants such as "Ceteris paribus, if parameter value p does (not) exceed a threshold value, the model shows behavior H." The example discussed in section 3 provides a case in point: it concerns a robust result of the PRT′ form, which may—but need not—be found to be of the RT′ form after further analysis.

Another, perhaps larger, gap—namely, an absolute distinction in roles rather than a gradual, context-sensitive difference in fulfillment of the same role—separates parameter and structural robustness analysis from the third type identified by Weisberg and Reisman. In this representational robustness analysis, members of the model family considered differ in more than

https://doi.org/10.1086/666061 Published online by Cambridge University Press

structural features: they do not share a "representational framework" (Weisberg and Reisman 2008, 120). We leave aside the question of how a general distinction between mathematical structure and representational framework can be made, especially on the first, 'broad-mathematical-framework' reading of low-level confirmation (see sec. 2). Instead, we note that, in general, extending the considered model family to the point where they do not share even their most basic mathematical structure makes robustness analysis unfit for providing warrant for either the models or any of their shared behavior. Once all prior restrictions are lifted, only completeness can assure that similarities are not the result of an inappropriately limited analysis. This limits robustness analysis to an explorative role, as discussed in section 2, or to a third, *model-constructive* role.

For rather than discovering that an implication *H*, known from models with a familiar representational framework, is shared by a model with a different framework, it may be used as a requirement for the new mathematically unrelated model that it entails *H*. Then, robustness analysis serves as a design specification for the new model. The example of representational robustness given by Weisberg and Reisman—investigating which agent-based models exhibit the Volterra property familiar from population models—serves as a case in point. Since the Volterra property and principle are empirically well-documented phenomena in ecology (Weisberg and Reisman 2008, 108), an agent-based model that does not reproduce the Volterra property would lead us to question the model, not the status (robustness, plausibility, truth) of the property. In line with this, Weisberg and Reisman report that one intermediate result of their modeling efforts does not exhibit the Volterra property because, under all parameter settings of the model, one or both species invariably goes extinct. They do not take this as evidence against the Volterra property or as a reason to reject the model on empirical grounds (e.g., by pointing out the actual coexistence of prey and predators). Instead, they conclude that they "need to find an IBM [individual-based model] that exhibits coexistence of the two species [for if not, the Volterra property cannot arise by definition]" (125). So they reject the model and continue their quest for a new model that can reproduce the Volterra property, and they finally succeed by adding a factor accommodating carrying capacity.

Before turning to our third cautionary claim, let us consider an objection to a principled distinction between the (possibly evaluative) role of parameter and structural robustness and the (necessarily nonevaluative) role of representational robustness. In some cases, one might argue, representational robustness could be evaluative, namely, where a finite (probably small) number of distinct model families is used in a scientific discipline and none is strictly preferable to all others. A case in point is nuclear physics, in which both shell models and liquid-drop models are used for predictive and

https://doi.org/10.1086/666061 Published online by Cambridge University Press

explanatory purposes. These models (or rather model families) are so distinct that they seem good candidates for having different representational frameworks, and there is no more encompassing theoretical framework that yields accurate predictions—both model families are needed. Suppose, for the moment, that there are no other credible models in nuclear physics. Then, studying to what extent shell models and liquid-drop models behave similarly identifies theorems that are robust and supported by current credibility expectations. Yet even in this case, there is little reward to be gained regarding both the results and the models: the current status of the models is that there is warrant to believe results of either model, whether or not they are entailed by the other, and that no model is accepted as predictively or explanatorily adequate for all purposes. Instead, it is accepted that one model is better suited for some purposes, and the other for different purposes. Finding that some behavior is shared is largely irrelevant, given the credentials of the two incompatible model families. Thus, even in cases in which there are restrictions on varying representational frameworks—and completeness is not sought—representational robustness has no evaluative power.[17]

Our third and final cautionary claim concerns the confluence of robustness analyses. Given that types of robustness presuppose different model families and may play different roles, there is no benefit to be gained from mixing types or specific results of analyses. "Taken together" (Weisberg and Reisman 2008, 108), the types do not play their roles more successfully than they do separately. Weisberg apparently acknowledges this when suggesting that the three types of robustness analysis form a hierarchy: one might vary parameter values, but this only makes sense against a background of fixed mathematical structure, and one might vary structural features, but only within a single representational framework. The converse also holds, however. If the representational framework is varied, for instance, when comparing the behavior of liquid-drop and shell models of atomic nuclei, the fine structure of both model families is not considered. Likewise, when structural features are varied, one typically does not consider the finer details of parameter value settings. Take, again, the example of coupled-oscillator models. When studying how models with and without friction behave, one does not simultaneously study the influence on model behavior of force constants. Consequently, for evaluative purposes, a theorem that is either parametrically or structurally robust is not 'half as robust' as a theorem that is

17. This allows us to comment on a recent argument that all economic modeling consists in robustness analysis (Kuorikoski, Lehtinen, and Marchionni 2010). All modeling practices are indeed likely to involve studying and comparing the behavior of models against the background of family credentials. We would caution, however, that this comparison may serve different purposes in scientific practice and that these purposes can be carefully distinguished as different roles for robustness analysis.

https://doi.org/10.1086/666061 Published online by Cambridge University Press

both. More strongly, the results of parameter and structural robustness can only be combined with those of representational robustness if they are all used in a nonevaluative role. There, the results of exploration do show some confluence, although parameter and representational robustness differ significantly in the scope of exploration. Finally, for model-constructive purposes, only one specific type is ever applied since the point is to design a model with different parameter settings/structure/representational framework and the same implication.

5. Conclusion. We have argued for three cautionary claims with regard to (Weisberg's reconstruction of) robustness analysis. Most important, we have pointed out that there are limitations to the evaluative power of robustness, arising from the need to make a prior choice of a credible model family, and we have put glosses on a unified and formal account of robustness, on the basis of an identification of several roles—evaluative, explorative, and model-constructive—that robustness analysis may play in modeling practices and on the context sensitivity of (the relevance of) these roles. We take these remarks as making explicit and clarifying some elements of Weisberg's reconstruction and as providing a richer account of the significance of robustness analysis. Whatever criticisms we have voiced are subsidiary to this constructive purpose.

REFERENCES

Bass, Frank M. 1969. "A New Product Growth Model for Consumer Durables." *Management Science* 15:215–27.

Boumans, Marcel. 1999. "Built-in Justification." In *Models as Mediators*, ed. Mary S. Morgan and Margaret Morrison, 66–96. Cambridge: Cambridge University Press.

Farrell, Christopher J. 1993. "A Theory of Technological Progress." *Technological Forecasting and Social Change* 44:161–78.

Fisher, J. C., and R. H. Pry. 1971. "A Simple Substitution Model of Technological Change." *Technological Forecasting and Social Change* 3:75–88.

Forber, Patrick. 2010. "Confirmation and Explanations How Possible." *Studies in History and Philosophy of Biological and Biomedical Sciences* 41:32–40.

Forster, Malcolm, and Elliott Sober. 1994. "How to Tell When Simpler, More Unified, or Less *Ad Hoc* Theories Will Provide More Accurate Predictions." *British Journal for the Philosophy of Science* 45:1–35.

Henrich, Joseph. 2001. "Cultural Transmission and the Diffusion of Innovations." *American Anthropology* 103:992–1013.

Kieseppä, I. A. 2001. "Statistical Model Selection Criteria and the Philosophical Problem of Underdetermination." *British Journal for the Philosophy of Science* 52:761–94.

Kuorikoski, Jaakko, Aki Lehtinen, and Caterina Marchionni. 2010. "Economic Modelling as Robustness Analysis." *British Journal for the Philosophy of Science* 61:541–67.

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

Lloyd, Elizabeth A. 2010. "Confirmation and Robustness of Climate Models." *Philosophy of Science* 77:971–84.

Meade, N., and T. Islam. 1998. "Technological Forecasting—Model Selection, Model Stability and Combining Models." *Management Science* 44:1115–30.

https://doi.org/10.1086/666061 Published online by Cambridge University Press

Michelfelder, R. A., and M. Morrin. 2005. "Overview of New Product Sales Forecasting Models." In *Intellectual Property*, ed. G. V. Smith and R. L. Parr, 817–28. Hoboken, NJ: Wiley.
Morris, Steven A., and David Pratt. 2003. "Analysis of the Lotka-Volterra Competition Equations as a Technological Substitution Model." *Technological Forecasting and Social Change* 70:103–33.
Orzack, Steven H., and Elliott Sober. 1993. "A Critical Assessment of Levins's 'The Strategy of Model Building in Population Biology' (1966)." *Quarterly Review of Biology* 68:533–46.
Porter, Alan L., A. Thomas Roper, Thomas W. Mason, Frederick A. Rossini, and Jerry Banks. 1991. *Forecasting and Management of Technology.* Hoboken, NJ: Wiley.
Saviotti, P. P., and G. S. Mani. 1995. "Competition, Variety and Technological Evolution." *Journal of Evolutionary Economics* 5:369–92.
Weisberg, Michael. 2006. "Robustness Analysis." *Philosophy of Science* 73:730–42.
———. 2007. "Who Is a Modeler?" *British Journal for the Philosophy of Science* 58:207–33.
Weisberg, Michael, and Kenneth Reisman. 2008. "The Robust Volterra Principle." *Philosophy of Science* 75:106–31.
Wimsatt, William C. 1981. "Robustness, Reliability, and Overdetermination." In *Scientific Inquiry and the Social Sciences*, ed. M. Brewer and B. Collins, 124–63. San Francisco: Jossey-Bass.
Woodward, Jim. 2006. "Some Varieties of Robustness." *Journal of Economic Methodology* 13:219–40.

https://doi.org/10.1086/666061 Published online by Cambridge University Press