Forthcoming in *The Routledge Handbook of Philosophy of Scientific Modeling*,
ed. by Natalia Carrillo, Tarja Knuuttila and Rami Koskinen, Routledge, 2023

Robustness analysis

Wybo Houkes, Dunja Šešelja and Krist Vaesen

Abstract. This chapter is devoted to robustness analysis, a common practice in modelling, where researchers vary features of a model and study the impact of changes on its behavior. After presenting the three most prominent types discussed in the philosophical literature, the chapter reviews the debate surrounding the epistemic role of this practice, focusing on the contested issue of its evidential import. The discussion highlights the multiple roles of robustness analysis, including the value of not establishing the robustness of a particular modeling result.

1. Introduction: what is robustness analysis?

In most modelling practices, researchers do more than *construct* and *manipulate* models. In order to draw conclusions on the phenomena that these models are taken to address, they also *vary* features of the model and study the impact of these changes on the model's behavior. These practices are found across disciplines and contexts of application and, in many of these, are known as *robustness analysis*.[1] Under this heading, we may find ecologists examining how changes in parameter settings affect the behavior of Lotka-Volterra equations, taken to represent interacting populations of organisms, physicists studying the impact of perturbation terms on Navier-Stokes equations that represent turbulence, and social scientists checking how Schelling models of segregation depend on particular relocation rules.

For philosophers of science, the main interest has been to understand why modelers engage in this practice, i.e., what is epistemically valuable in robustness analysis (henceforth: RA). As James Woodward put it in the context of economic modelling, the aim is to understand whether and, if so, why "robustness (of inferences, measurements, models, phenomena and relationships discovered in empirical investigations etc.) is a Good Thing" (Woodward 2006: 219). Robustness here stands for the stability of these inferences / measurements / models / phenomena under perturbations affecting the broader context or a system they belong to. While robustness in a broader sense has been used to capture different notions of stability, we focus on robustness of results obtained by means of scientific models and RA as a method of examining this property.[2]

The most prominent explanation, which arguably started the current discussion, is found in Richard Levins' work. Levins describes RA as a powerful strategy available to modelers like him:

> "... 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*" (Levins 1968: 423; emphasis added)

Levins' description makes evident the potential value of RA: it would allow modelers to derive true claims from models that are in important respects inaccurate or (over)simplified.

In the extreme case, genuine insights into complex real-world systems could be gained by studying only a variety of highly unrealistic, minimal or 'toy' models. Although this would clearly be 'a Good Thing', philosophers have understandably suspected that it is too good to be true.

In this chapter, we review the ensuing debate. In the philosophy of science, a key role was played by William Wimsatt (1981), who identified the three central elements of RA that philosophers are still mainly concerned with: its *core definition* and *varieties*; its epistemic *value*; and the *conditions* under which it realizes this value. We briefly review each, also to set the stage for this paper.

Regarding the *central definition*, Wimsatt notes that a broad variety of practices can be gathered under the heading of 'robustness analysis'. This includes checking which implications of models remain the same under change to those models, but also practices such as triangulation, which check whether observational results remain the same under change of method. In all of these, the aim is to determine whether something is 'robust', where:

> "*X* is robust = *X* remains invariant under a multiplicity of (at least partially) independent derivations" (Soler et al. 2012: 3, paraphrasing Wimsatt 1981)

Wimsatt's reasons for discussing the practices under the same heading refer directly to RA's most contentious features: its overriding purpose or epistemic value, as well as the conditions for realizing this purpose or value – the reason for engaging in these practices, and their proper implementation. For both, Wimsatt extends and partly specifies Levins' characterizations. Regarding purpose, "[a]ll the variants and use of robustness have a common theme in … distinguishing … which is regarded as ontologically and epistemologically trustworthy and valuable from that which is unreliable, ungeneralizable, worthless, and fleeting" (Wimsatt 1981/2012, p.63). More extensively than Levins, Wimsatt identifies necessary conditions for realizing this, as well as a risk of engaging in RA:

> "[a]ll these procedures require at least some partial *independence* of the various processes across which invariance is shown. And each of them is subject to a kind of systematic error leading to a kind of *illusory robustness* when we are led, on less than definitive evidence, to presume independence" (ibid., 64; emphasis in original)

As the latter part of the quote makes clear – more so than Levins' much-quoted claim – there is a risk to engaging in RA. Because of this systematic error, which Wimsatt claims is intrinsic to the practice, it makes sense to investigate which, if any, of the varieties of RA meet which conditions for successfully realizing the envisaged purpose.

In this chapter, we review this debate and its results so far. We do so by focusing, like most philosophers of science, on the role of RA in the testing of model-derived theorems for an epistemic (rather than ontological) purpose. Some in the debate defend that RA can realize the purpose envisaged by Levins and Wimsatt – albeit only in some forms and under strict conditions and qualifications. Others reject this, mainly by problematizing Wimsatt's condition of independence; they submit that any robustness will, on closer inspection, turn out to be illusory for evidential purposes. However, critical authors have identified alternative epistemic purposes of robustness analysis. Interestingly, in some cases *negative* results (i.e., the 'fragility' of an implication) can be equally or even more valuable than positive results. So, where Woodward's framing suggests that lack of robustness is a Bad Thing, modelling practice does not always conform, and modelers might have many options to manage the risk of 'illusory robustness' mentioned by Wimsatt and emphasized by many philosophers.

We start by introducing some terminology and reviewing the three most prominent types of RA that have been distinguished by philosophers of science (Section 2). In Section 3, we turn from types of RA to the various roles or epistemic functions of it, focusing on the contested issue of its evidential import. Section 4 concludes the chapter.

2. Different types of robustness analysis

Before presenting the most prominent types of RA discussed in the philosophical literature, we define some key terms. In the literature, 'robustness analysis' refers to any practice of varying aspects of the model and studying which implications remain invariant; and 'robustness' refers to any invariance revealed. In RA, relevant aspects of a model are changed, and it is established whether particular implications of this model are invariant under those changes. Implications that are invariant to a relevant degree are called 'robust'; and we refer to the models that share the implication as the 'robustness set' for the implication. Some authors, following Levins (1968), take the result of (successful) robustness analysis to be a *robust theorem* rather than an implication. This requires an additional analytical step, to identify the minimal features shared by members of the robustness set that entail the invariant implication (Weisberg 2006; Weisberg and Reisman 2008).[3]

Robustness analysis is a systematic way or strategy of identifying a robustness set: it starts from a model *M*, varies it in some respect, and checks whether some relevant implication *p* is conserved; here, *M* and *p* may be called the 'targets'. RA is thus a generative method, rather than a merely comparative one, in which one would search for some arbitrary alternative model that has a sufficiently similar implication. Finding out, for instance, which (if any) implications are shared by magnetohydrodynamic models of fusion plasmas and Schelling's checkerboard models of segregation would not be called 'robustness analysis', if it is a meaningful scientific practice at all.

Following the relevant literature in the philosophy of science, one can distinguish three prominent types of RA. Each concerns a different way of generating the robustness set, i.e., each type primarily indicates in which respect a target model is changed to determine the effects on a target implication. In the literature, different typologies as well as nomenclatures can be found.[4] We follow Weisberg and Reisman (2008) both in the nomenclature and in distinguishing these three types of RA.

In *parameter* RA, it is checked whether some implication of a model and its auxiliary assumptions is robust to the extent that the implication holds over different parameter settings. Thus, the robustness set is generated by varying the parameters of a target model over some interval. Take, for instance, Schelling's model of social segregation. The model was designed to examine factors concerning individual preferences that lead two groups within a society to segregate. Schelling approached this question in terms of an abstract model: by randomly placing members of two groups of an equal size on a checkerboard, he examined how the population changes if we assume that individuals have a specific preference about the composition of their neighborhood. One striking result of this model is that, even when agents prefer as little as one third of their neighborhood to consist of members of their own group, the society ends up clustered in homogenous neighborhoods: there is '*de facto* segregation with mild in-group preferences'. To examine the parameter robustness of this implication of the Schelling model, we can test whether similar *de facto* segregation obtains once we change the size of the population, the size of the checkerboard, and so forth.

Some authors have called parameter RA 'sensitivity analysis' (e.g., Raerinne 2013, Gräbner 2018). In some disciplines, such as many forms of economic modelling, practices under the latter name indeed largely match what we described immediately above (i.e., checking to what extent implications are conserved under varying parameter settings). However, in some contexts and disciplines, 'sensitivity analysis' refers to a broader set of practices, in which one investigates how the output of a model changes under variations in input parameters (see, e.g., Saltelli et al. 2008 for an overview of techniques). Here, modelers are not specifically interested in output invariance, i.e., robustness; rather, they seek a more general understanding of the relations between a model's input and output, e.g., to identify which input variables most strongly affect output ('importance assessment'; Saltelli 2002).

Structural RA pertains to structural features of the target model, in particular its central assumptions.[5] In this case the modeler aims to find out which parts of the model's structure govern an implication. Such an analysis can take two forms. First, the modeler might remove or relax certain existing assumptions. Second, the modeler might add assumptions or replace existing ones. In either of these ways, modelers may find out which assumptions are genuine difference-makers with regard to the implication. In particular, structural robustness may test the implication's dependence on what Kuorikoski et al. (2010) distinguish as 'tractability assumptions' and 'substantial assumptions'. The former are mathematical formulations allowing for an easier or more efficient solution of the represented problem. Such assumptions usually have no clear causal interpretation and/or are highly unrealistic. They are a 'necessary evil', intended to facilitate derivations or even to make them feasible at all. Substantial assumptions, on the other hand, are empirically informed and they serve to identify the causal structure of the target phenomenon.[6] While tractability assumptions may impact the formal representation of substantial assumptions, substantial assumptions may impact the tractability of the model. Such dependencies may restrict the scope of structural RA for some implications and assumptions: for lack of tractable results, it may be impossible to determine the effects of target implications for some relevant changes.[7]

For instance, network epistemology models, which study the impact of social networks on the production of knowledge, usually represent the structure of information flow in terms of directed graphs, with nodes standing for agents and edges between them for communication channels. This allows for the representation of communities that have varying degrees of connectivity, that is, a varying degree of information flow. Structural RA can, on the one hand, be used to examine whether changing such a tractable representation of information flow impacts the result of the model. For instance, Borg et al. (2017) examine whether the results of their model remain stable once a network in which the probability that an agent shares information with others is a parameter of the model, is replaced with networks that have stable links between agents. On the other hand, structural RA can be used to study the impact of different substantial assumptions, such as those that underpin the representation of learning. For example, if agents stand for scientists who are trying to identify the better of two available theories, we can represent their research in different ways. We could, for instance, assume that scientists have 'inertia' towards their preferred theory in the sense that they do not immediately abandon it even if they learn from others that the alternative theory appears to be better. As such behavior of scientists may be more characteristic of some contexts of inquiry rather than others, the assumption is an empirical issue. Frey and Šešelja (2020) use structural RA to examine the impact of adding such inertia to the process of scientific research in Zollman's (2010) network epistemology model to specify the context of learning to which the results of the model apply.

Representational RA goes beyond structural RA in varying the representational framework, modelling technique, or modelling medium. The aim here is to determine the extent to which the target model's specific representational framework or implementation makes a difference with respect to its implications. For instance, the Volterra principle was originally derived from a set of differential equations, which describe predation at the population level. Using representational RA, one may study whether the principle also holds if the predatory system is represented in terms of individuals and their individual-level properties. Indeed, Weisberg and Reisman (2008) present a set of such agent-based models and find that they too produce the Volterra principle. From this, the authors conclude that the principle is robust across at least two representational frameworks. Another example is evolutionary game-theoretic modelling, which is based either on mathematical analytical frameworks or on computational frameworks such as agent-based models (ABMs). As de Marchi and Page (2009) argue, ABMs allow for the representation of features that may be impossible to represent in analytical models due to tractability constraints. Again, implications that are shared by ABMs and analytical models may be called (representationally) robust; here, one may conclude specifically that these implications are not artefacts of the constraints inherent to analytical frameworks. As such, representational RA can, like structural RA, serve to study the impact of certain tractability assumptions in the models. Finally, modelers may vary the medium in which models are realized or implemented: Knuuttila and Loettgers (2021) discuss how, in synthetic biology, a particular network design (the Repressilator Model) was implemented in multiple media to test whether it produced robust oscillations in genetic networks.

Intuitively, the change made in representational RA is 'larger' than the one in structural RA: it concerns the very formal modelling technique rather than a particular tractability assumption made in implementing a technique. The robustness set in representational RA thus also consists of models that hold a stronger (intuitive) claim to being independent, since they are not constructed with the same technique or, more broadly, epistemic means. In light of Levins' claim, this would seem to make *positive results* of representational RA more valuable than those of structural or parameter RA. Admittedly, examples of such positive results are also difficult to find, whereas variations of parameter settings and structural features are part and parcel of modelling practice. This, however, may only underline how valuable representationally robust implications are if they can be obtained (cf. Houkes & Vaesen 2012, Lisciandra 2017).

The main purpose of representational RA is perhaps in negative findings: failing to replicate a result with a different framework may help to identify a set of difference-making assumptions in the original model, which may otherwise remain overlooked. For instance, in the above-mentioned field of network epistemology, Borg et al. (2018) use an agent-based model (ABM) based on argumentation dynamics to examine the robustness of results previously obtained with an ABM employing a Bayesian framework based on bandit models (Zollman 2010). While Zollman's results are representationally robust with respect to a number of ABMs employing the epistemic landscape framework (e.g., Lazer and Friedman 2007, Grim et al. 2013), Borg et al. fail to reproduce the same findings. In light of this, Borg et al. identify assumptions in their model, absent from the previous ones, which are responsible for this outcome. This in turn helps to specify the context of learning to which previous results apply.

3. Epistemic roles

Philosophers of science have discussed various epistemic roles that robustness analysis can play. Most of the discussion has focused on the question under which conditions (if any) this role can be evidential – roughly, when modelers have indeed found a 'truth at the intersection of independent lies'; and slightly less roughly, whether positive results of RA should increase one's credence in the truth of some hypothesis. Insofar as other epistemic roles have been discussed, this was mainly to identify an alternative, which would make sense of modelers' engaging in RA even when it cannot play an evidential role. In this section, we first outline the main arguments regarding the evidential role of RA, and then review some of the alternative roles that have been identified.

3.1. Does robustness analysis have evidential value?

Levins' original claim can be read in a strong way: showing that an implication is robust provides evidence for regarding this claim as true, i.e., by studying whether a set of models behaves similarly, one can learn something about the world. Furthermore, Levins suggests that RA could play this strong evidential role regardless of any observational evidence for this implication or a robust theorem. This would make RA especially valuable if it is difficult or impossible to validate a model or its implications in another way, e.g., by successful prediction. Such an epistemic situation obtains in many modelling contexts across research fields, e.g., in economics, evolutionary biology, climate science and computational philosophy. Consequently, many contributions to the debate draw on one or more of these contexts to illustrate their general claims – positive or negative – about the role of RA.

It is broadly acknowledged (e.g., Cartwright 1991; Orzack and Sober 1993; Sugden 2000), that RA does not have the strong, complementary evidential role suggested by Levins' dictum – or at least, that the conditions for RA playing this role are so strict that this cannot reasonably explain the widespread use of the practice. To see why, take an extended Schelling model in which agents' behavior is governed by their 'range of vision' R over the grid (with R a natural number), rather than only their immediate neighbors (corresponding to R = 1) (Laurie and Jaggi 2003). Suppose that some interesting implication p holds for all ranges *R*, i.e., that *p* is parameter-robust with respect to *R*. Then, we may conclude that *p* is true for actual urban areas – or other target systems to which Schelling models are applied – only if a modeler has reason to believe that the correct model of the target system may be found in this robustness set, consisting of models in which $R \in [1, m]$, where *m* is the measure of the grid length. If the modeler does not know whether this is the case, let alone if she has reason to think that all members of the robustness set are unrealistic in some relevant respect, *R*-robustness alone does not have sufficient evidential impact to warrant accepting the target implication. In Levins' terms, something has been found at the intersection of lies, but it cannot be said to be a truth.

In response, it could be pointed out that this analysis ignores one important aspect of Levins' statement: the models in the robustness set need to be *independent*. Recall that according to Wimsatt, failure of independence produces illusory robustness, and that the models in the set need to have "at least some *partial* independence" (see Section 1). Only if the models are mutually independent can RA play a role similar to triangulation, making it less likely that the implication is false.

A well-established line of argumentation shows the difficulties in spelling out a suitable notion of independence. As Orzack and Sober (1993) point out, competing models of the same phenomenon cannot be *logically* independent, since the truth of one implies the falsity of all the others. Models in robustness sets tend to be competing. Take, for instance, our case from above: at most one value of *R* can be descriptively adequate for a given urban area. The models in a robustness set are not *statistically* or *probabilistically* independent, in the sense that a certain result following from one model has no bearing on the probability that the same result will be detected by the other model (cf. Schupbach 2018, who also discusses other notions of independence in this context). However, when doing RA, modelers do not review models that are independent in this way. Reviewing whether target implications still hold under changes of parameter settings requires holding fixed a model's structural assumptions. While the latter assumptions may be relaxed or changed (in structural RA), deriving implications typically requires holding fixed the model's tractability assumptions. Finally, checking whether implications hold under changes in tractability assumptions requires holding fixed substantial assumptions (including structural assumptions and those concerning parameter values). Even if this is done via representational RA, the chosen representational frameworks need to have the core substantial assumptions in common. Therefore, in a crucial sense, the models in a robustness set must share some of their assumptions. As a result, robustness might still only reflect commonalities of the models and/or the representational frameworks (cf. Odenbaugh and Alexandrova 2011: 763). In Orzack and Sober's words, there is always the possibility that "robustness simply reflects something common among the frameworks and not something about the world those frameworks seek to describe" (1993: 539). Phrased more negatively, using Wimsatt's terms, no notion of 'partial' independence seems available which would dispel the suspicion that robustness might be illusory and confer evidential value on RA.

A recent, powerful defense of the evidential role of RA grants the validity of this critical argument, but submits that it largely misses the point of how RA can be and is used in modelling practice. According to Kuorikoski et al. (2010, 2012), epistemically impactful RA does not feature just any change to a model (let alone every possible change); rather, it focuses on specific assumptions to show that a target implication does not crucially depend on them. While this does not amount to empirical confirmation of the implication, it should also not be dismissed as epistemically futile. According to the authors, the primary value of RA lies in making our inferences *more reliable* and *increasing our confidence* in them by showing that they do not depend on problematic modelling assumptions. Since RA serves to identify assumptions that the result of the model depends on, if such assumptions are problematic, this will lower our confidence in the given inference. However, if the result appears to depend mainly on plausible substantial assumptions, we should have more confidence in its validity than prior to conducting the RA. Importantly, for RA to play such an evidential role, the substantial modelling assumptions need to be 'reasonably realistic'. In other words, RA can increase our confidence in the given inference only in combination with empirical evidence supporting the assumptions of the model.[8] Moreover, for RA to have this effect, there should be no reason, prior to RA, to think that differences in tractability assumptions of the studied models "have a similar mathematical and empirically interpretable impact on the modelling result" (Kuorikoski et al. 2012: 898). In Levins' terms, RA requires independence of the specific *lies* inherent to each model in the set; then, a robust result might still not be true, but it is at least not an artefact of one specific lie.

This debate on the evidential role of RA has revealed that this role is tightly connected to empirical underpinnings of the studied models. For models with realistic substantial assumptions, RA can serve to insulate (some) implications from (some) specific lies, such as particular parameter settings, auxiliary assumptions, idealizations, or even tractability assumptions.[9] It might also provide indirect confirmation if the robustness set of the implications consist of models that have other confirmed results (Lehtinen 2018). Defenders of this evidential value admit, however, that robustness could always prove to be illusory, because implications could be the result of shared and unquestioned assumptions within or even across modelling frameworks.[10] Use of a large number of such frameworks may alleviate this worry to some extent, since they are unlikely to all share such assumptions. Whether or not they do, however, remains an empirical question; there is no strength in numbers here per se.

3.2. Which other epistemic roles can robustness analysis play?

An interesting side-effect of the debate on the evidential role of RA has been the identification of various alternative purposes that RA can and does serve in modelling practices. The reason is, of course, that if RA cannot or hardly ever increases our credence in hypotheses, it becomes all the more puzzling 'what modelers get out of it': why is the practice so widespread if positive robustness checks do not give (additional) reasons to believe that particular modelling results are true? Even if one would assign an evidential role to RA, alternative roles could be used as supplementary reasons to engage in the practice. Here, we briefly describe several alternatives that have been identified.

I. Discovery of causal structure.
Even those who are not convinced that RA might have evidential value often subscribe to its usefulness in generating causal hypotheses. Specifically, RA allows exploration of the implications of substantial assumptions, together with varying parameter settings, tractability assumptions, auxiliary assumptions, etc. If such substantial assumptions identify the causal structure of a phenomenon, these explorations allow statements about the conditions in the

model-world under which the causal mechanism holds. In this way RA allows for the formulation of more precise causal hypotheses,[11] or to identify the common causal mechanism in a family of models, rather than providing evidence for any implications. Thus, Knuuttila and Loettgers (2011) distinguish 'causal isolation' RA from the 'independent determination' RA on which most of the philosophical literature has focused. In this epistemic role, RA can also help to formulate pursuit-worthy hypotheses. It does so by providing 'inquisitive reasons' (Fleisher 2022), which are reasons that concern promoting successful inquiry (such as showing that a hypothesis is testable, that it is based on a heuristic analogy, etc.). By identifying specific conditions under which the given causal mechanism holds in the model-world, RA helps to delineate the application domain in which the causal hypothesis should be further pursued in terms of empirical studies.

II. Deepened causal understanding.
Relatedly, and perhaps a bit more distinctively, RA might help to develop and deepen our causal understanding of real-world systems and phenomena. It may do so by presenting a way in which to vary systematically – albeit through their representation in substantial assumptions, and heavily mediated by tractability assumptions and other auxiliaries – the factors that could be causally responsible for certain system behavior. All forms of RA would appear to be useful in this respect. Parameter RA helps to study the range under and extent to which factors cause behavior (e.g., how the 'range of vision' influences segregation in Schelling models; Laurie and Jaeggi 2003). Structural RA contributes to developing more sophisticated causal understanding, because it allows studying the effects of adding or removing factors as well as possible confounders and mediators. Finally, representational RA allows studying alternative or supplementary causal mechanisms, perhaps at different levels of organization (e.g., population-level versus individual).[12]

III. Elimination of (alternative) potential explanations
As a complement to the previous role, RA might serve an eliminative role in explanatory reasoning, as argued by Schupbach (2018). Suppose that we have a model that has some empirically validated implications and we are trying to explain why the model gives this result. Then, studying how these implications of the model vary under changes to the model may serve to rule out competing possible explanations of this kind. For instance, in the case of the Volterra principle, this means ruling out various explanations which stipulate that the result is due to idealizing and simplifying assumptions in the model. Specifically, if such competing alternatives entail that implications fail to hold under particular changes, this provides a way of discriminating between them and the target explanation. In case of the above example this means that RA can help to discriminate between two explanations: that the model accurately represents the given predator-prey dynamics and therefore continues to behave in accordance to the Volterra principle if we relax certain unrealistic assumptions; or that the result is due to the given unrealistic assumption (so that, once this assumption is removed, we should fail to observe the same output). RA could thus amount to a strategy of systematically and incrementally generating such explanatorily discriminating means.[13]

IV. Calibration of alternative modelling techniques
RA may have a role in *constructing* models rather than in studying and evaluating their implications. This is most straightforwardly illustrated with representational RA. When developing a modelling technique as an alternative to existing approaches, some implications may be used to calibrate or even test the alternative: only if those implications

can be replicated, the alternative will be considered. Houkes and Vaesen (2012: 361) argue that this applies to Weisberg and Reisman's agent-based alternative to Lotka-Volterra models: an alternative that does not display the Volterra property (i.e., the desired implication) is discarded in favor of another, more sophisticated agent-based model. Structural RA might play out similarly, for instance if changing structural features of a model only reproduces desired results under specific parameter settings or with additional auxiliary assumptions. This calibrative role of RA is, in many ways, complementary to the eliminative role discussed above. Clearly, it has no bearing on one's credence in any hypothesis, since there is not even the semblance of independence; thus, if one adopts Levin's and Wimsatt's characterization of RA, this practice may be taken as a degenerate case of the practice.

4. Conclusion

Robustness analysis is commonly used in modelling practices as the method of examining the stability of results under various perturbations of features of the model. In light of this, philosophers of science have inquired which kinds of RA there are, and what exactly their epistemic function is. In this chapter, we have reviewed this debate. We started by defining key terms and distinguishing between parameter RA, structural RA and representational RA. While each kind of RA can increase our understanding of the studied models, philosophers have debated whether any of them can have an evidential, confirmatory value in the sense that a robust modelling result can be considered true of real-world phenomena. Even though there is general consensus in the literature that RA on its own does not provide an evidential import of that kind, different proposals of its alternative epistemic functions have been put forward. As our discussion shows, RA can help to improve not only our understanding of the inner functioning of models, but also our causal and explanatory insights obtained by them. Yet, for RA to play such a role it has to be combined with empirical methods, on the basis of which the model and its results can be empirically embedded in the first place. Whether and to which extent this is possible remains a challenge for each domain of modelling, especially for those that employ either highly idealized, theoretical models or highly complex but difficult-to-validate models. Moreover, which types of RA are most epistemically useful in such cases – and whether *negative* results of RA can be as much of a Good Thing as positive results – is another question that may vary from one modelling context to another.

References

Borg, AnneMarie, Daniel Frey, Dunja Šešelja, and Christian Straßer. 2017. "Examining Network Effects in an Argumentative Agent-Based Model of Scientific Inquiry." In *International workshop on logic, rationality and interaction*, ed. Alexandru Baltag, Jery Seligman and Tomoyuki Yamada, 391–406. Lecture Notes in Computer Science 10455. New York: Springer.

Borg, AnneMarie, Daniel Frey, Dunja Šešelja, and Christian Straßer. 2018. "Epistemic Effects of Scientific Interaction: Approaching the Question with an Argumentative Agent-Based Model." *Historical Social Research* 43 (1): 285–309.

Calcott, Brett. 2011. "Wimsatt and the Robustness Family: Review of Wimsatt's *Re-engineering Philosophy for Limited Beings.*" *Biology & Philosophy* 26: 281–93.

Cartwright, Nancy. 1991. "Replicability, Reproducibility, and Robustness." *History of Political Economy* 23: 143–55.

Casini, Lorenzo, and Jürgen Landes. 2022. "Confirmation by Robustness Analysis: A Bayesian Account." *Erkenntnis*. DOI: 10.1007/s10670-022-00537-7

De Marchi, Scott, and Scott E. Page. 2009. "Agent-Based Modeling." In *The Oxford Handbook of Political Methodology*, ed. Janet M. Box-Steffensmeier, Henry E. Brady, and David Collier, 71–94. Oxford: Oxford University Press.

Eronen, Markus I. 2015. "Robustness and Reality." *Synthese* 192 (12): 3961–77.

Fleischer, Will. 2022. "Pursuit and Inquisitive Reasons." *Studies in History and Philosophy of Science* 94: 17–30

Forber, Patrick. 2010. "Confirmation and Explaining How Possible." *Studies in History and Philosophy of Science Part C* 41: 32–40.

Frey, Daniel, and Dunja Šešelja. 2020. "Robustness and Idealization in Agent-Based Models of Scientific Interaction." *British Journal for the Philosophy of Science* 71 (4): 1411–37.

Fuller, Gareth P., and Armin W. Schulz. 2021. "Idealizations and Partitions: A Defense of Robustness Analysis." *European Journal for Philosophy of Science* 11: 1-15.

Gräbner, Claudius. 2018. "How to Relate Models to Reality? An Epistemological Framework for the Validation and Verification of Computational Models." *Journal of Artificial Societies and Social Simulation* 21 (3): 8.

Grim, Patrick, Daniel J. Singer, Steven Fisher, Aaron Bramson, William J. Berger, Christopher Reade, Carissa Flocken, and Adam Sales. 2013. "Scientific Networks on Data Landscapes: Question Difficulty, Epistemic Success, and Convergence." *Episteme* 10 (4): 441–64

Houkes, Wybo, and Krist Vaesen. 2012. "Robust! Handle with care." *Philosophy of Science* 79 (3): 345–64.

Knuuttila, Tarja, and Andrea Loettgers. 2011. "Causal Isolation Robustness Analysis: The Combinatorial Strategy of Circadian Clock Research." *Biology & Philosophy* 26: 773–91.

Knuuttila, Tarja, and Andrea Loettgers. 2021. "Biological Control Variously Materialized: Modeling, Experimentation and Exploration in Multiple Media." *Perspectives on Science* 29 (4): 468–92.

Kuhlmann, Meinard. 2021. "On the Exploratory Function of Agent-Based Modeling." *Perspectives on Science* 29 (4): 510–36.

Kuorikoski, Jaakko, Aki Lehtinen, and Caterina Marchionni. 2010. "Economic Modelling as Robustness Analysis." *British Journal for the Philosophy of Science* 61 (3): 541–67.

Kuorikoski, Jaakko, Aki Lehtinen, and Caterina Marchionni. 2012. "Robustness Analysis Disclaimer: Please Read the Manual before Use!" *Biology & Philosophy* 27 (6): 891–902.

Laurie, Alexander J., and Narendra K. Jaggi. 2003 "Role of 'Vision' in Neighbourhood Racial Segregation: A variant of the Schelling Checkerboard Model." *Urban Studies* 40 (13): 2687–704.

Lazer, David, and Allan Friedman. 2007. "The Network Structure of Exploration and Exploitation." *Administrative Science Quarterly* 52 (4): 667–94.

Lehtinen, Aki. 2018. "Derivational Robustness and Indirect Confirmation." *Erkenntnis* 83 (3): 539–76.

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

Lisciandra, Chiara. 2017. "Robustness Analysis and Tractability in Modeling." *European Journal for Philosophy of Science* 7 (1): 79–95.

Lloyd, Elisabeth A. 2010. "Confirmation and Robustness of Climate Models." *Philosophy of Science* 77 (5): 971–84.

Mäki, Uskali. 1994. "'Isolation, Idealization, and Truth in Economics." In *Idealization VI: Idealization in Economics*, ed. Bert Hamminga and Neil B. De Marchi, 147–68. Amsterdam: Rodopi.

Odenbaugh, Jay, Anna Alexandrova. 2011. "Buyer Beware: Robustness Analyses in Economics and Biology." *Biology & Philosophy* 26 (5): 757–71.

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 (4): 533–46.

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

Paternotte, Cédric, and Jonathan Grose. 2017. "Robustness in Evolutionary Explanations: A Positive Account." *Biology & Philosophy* 32 (1): 73–96.

Raerinne, Jani. 2013. "Robustness and Sensitivity of Biological Models." *Philosophical Studies* 166 (2): 285–303.

Saltelli, Andrea. 2002. "Sensitivity Analysis for Importance Assessment." *Risk Analysis* 22 (3): 579–90.

Saltelli, Andrea, Marco Ratto, Terry Andres, Francesca Campolongo, Jessica Cariboni, Debora Gatelli, Michaela Saisana, and Stefano Tarantola. 2008. *Global Sensitivity Analysis: The Primer*. Hoboken, NJ: John Wiley & Sons.

Schupbach, Jonah N. 2018. "Robustness Analysis as Explanatory Reasoning." *The British Journal for the Philosophy of Science* 69 (1): 275–300.

Šešelja, Dunja. 2021. "Exploring Scientific Inquiry via Agent-Based Modelling." *Perspectives on Science* 29 (4): 537–57.

Šešelja, Dunja, and Christian Straßer. 2014. "Epistemic Justification in the Context of Pursuit: A Coherentist Approach." *Synthese* 191 (13): 3111–41.

Soler, Léna, Emiliano Trizio, Thomas Nickles, and William C. Wimsatt. Eds. 2012. *Characterizing the Robustness of Science: After the Practice Turn in Philosophy of Science*. Boston Studies in the Philosophy of Science 292. Dordrecht: Springer.

Stegenga, Jacob, and Tarun Menon. 2017. “Robustness and Independent Evidence.” *Philosophy of Science* 84 (3): 414–35.

Sugden, Robert. 2000. “Credible Worlds: The Status of Theoretical Models in Economics.” *Journal of Economic Methodology* 7: 169–201.

Weisberg, Michael. 2006. “Robustness Analysis.” *Philosophy of Science* 73: 730–42.

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 in the Social Sciences*, ed. Marilynn B. Brewer and Barry E. Collins, 123–62. San Francisco: Jossey-Bass. Reprinted in: (Soler et al. 2012), pp. 61–87.

Woodward, James. 2006. “Some Varieties of Robustness.” *Journal of Economic Methodology* 13 (2): 219–40.

Zollman, Kevin J.S. 2010. “The Epistemic Benefits of Transient Diversity.” *Erkenntnis* 72 (1): 17–35.

---

[1] See, e.g., Soler et al. (2012) for discussions of robustness analysis in various contexts of application.

[2] This means that we leave out other forms of robustness analysis, which would fit under Wimsatt’s more encompassing ‘multiple-determination’ heading. For instance, scholars have written about evidence robustly corroborating theories (Eronen 2015; Calcott 2011), about phenomena being robustly present in different contexts (Calcott 2011), or about robustness of scientific knowledge in a given domain (Šešelja & Straßer 2014).

[3] The same goes for understanding robustness analysis in terms of *robustness arguments*, e.g., Stegenga and Menon (2017), in which the set of statements in our scheme are the premises for the conclusion that *p* is more likely to be true.

[4] For instance: many authors follow Woodward (2006) in referring to parameter and structural robustness as ‘derivational robustness’; Kuhlmann (2021) calls representational robustness ‘multiple-model robustness’; etc.

[5] We prefer the term ‘structural' to ‘derivational’ RA since, similar to ‘parameter’ RA, it indicates the aspect of a model that is varied during the generative process of analysis.

[6] Kuorikoski et al. (2010) also distinguish ‘Galilean assumptions’ which are idealizations used to isolate the purported causal mechanism from all other interfering factors (see also, e.g., Mäki 1994).

[7] While Kuorikoski et al. (2010) consider derivational RA as an RA with respect to tractability assumptions, Raerinne (2013) introduces RA with respect to substantial assumptions as ‘sufficient parameter RA’ since different parameter values could be based on different substantial assumptions in the model.

[8] In a similar defense of RA, Michael Weisberg (2006) refers to the “low-level confirmation” of central modelling assumptions. Houkes and Vaesen (2012) identify some complications in this account. See Lloyd (2009) for an application of evidential RA to climate models based on Weisberg’s account, and Parker (2011) and Justus (2012) for a discussion of complications.

[9] “Robustness analysis is about coping with unavoidable falsity rather than finding *the* truth” (Kuorikoski et al. 2012: 899, emphasis in original).

[10] Schupbach (2018; Section 2) provides an in-depth review of other attempts to coin out the evidential value of RA. Also see Fuller and Schulz (2021) and Casini and Landes (2022).

[11] One way to develop this idea is in terms of open formulae - templates for formulating hypotheses that should then be empirically examined (Odenbaugh and Alexandrova 2011: 769).

[12] Paternotte and Grose (2017) discuss this and other explanatory roles of RA, focusing on evolutionary biology.

[13] Schupbach (2018; Section 3.2) reconstructs this role of RA so that it can have evidential value (e.g., with regard to mutually exclusive competing explanations). We discuss it as an alternative role

---

here since identifying this eliminative role does not seem to depend strictly on this reconstruction; Forber (2010), for instance, identifies a similar role for RA prior to empirical testing.