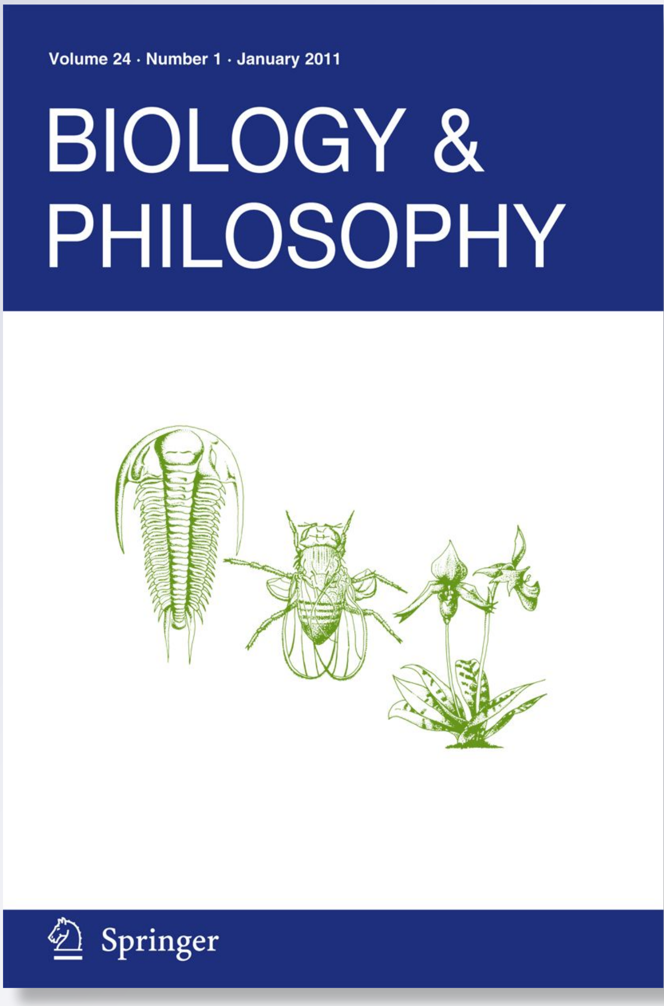
Buyer beware: robustness analyses in economics and biology

Jay Odenbaugh & Anna Alexandrova

Biology & Philosophy

ISSN 0169-3867

Biol Philos
DOI 10.1007/s10539-011-9278-y



Biol Philos
DOI 10.1007/s10539-011-9278-y

Buyer beware: robustness analyses in economics and biology

Jay Odenbaugh · Anna Alexandrova

Abstract Theoretical biology and economics are remarkably similar in their reliance on mathematical models, which attempt to represent real world systems using many idealized assumptions. They are also similar in placing a great emphasis on derivational robustness of modeling results. Recently philosophers of biology and economics have argued that robustness analysis can be a method for confirmation of claims about causal mechanisms, despite the significant reliance of these models on patently false assumptions. We argue that the power of robustness analysis has been greatly exaggerated. It is best regarded as a method of discovery rather than confirmation.

Keywords Model · Idealization · Robustness · Economics · Ecology · Game theory

Introduction

Theoretical biology and economics are remarkably similar in their reliance on mathematical models. These models are typically representations of familiar-sounding entities in familiar-sounding environments: prey among predators, agents in auctions, governments in negotiations, etc. Objects and properties that furnish the familiar world around us, or "commonsensibles" (Maki 1998), populate these

The authors are jointly and equally responsible for the content of the paper.

J. Odenbaugh
Department of Philosophy, Lewis and Clark College, Portland, OR, USA
e-mail: jay@lclark.edu

A. Alexandrova (✉)
Department of Philosophy, University of Missouri-St. Louis, St. Louis, MO, USA
e-mail: alexandrovaa@umsl.edu

models. But in addition, these models also use assumptions that heavily idealize the behavior and the environment of these entities, endowing them with infinite size, random mating, perfect reasoning, logical omniscience, and many more patently false yet apparently useful features.

A further striking similarity between the disciplines is that they prize robustness of model results. Robustness is here understood as "derivational robustness", to use James Woodward's term, i.e., the independence of the consequences of a model from particular assumptions used to derive these consequences (Woodward 2006). A modeler derives an interesting result from one class of assumptions, publishes it and then they or others roll up their sleeves and get busy checking whether this result would still hold under different assumptions. For example, what other shapes of utility functions, or other rules of the game, or other available information, or other functional responses imply or fail to imply the same prediction? If such alternative assumptions are found, then the original idealization is shown not to be essential and the prediction is shown to be, to some extent, robust. To produce robust theoretical results is, arguably, one of the fundamental goals of theoretical economists and biologists, and to evaluate various modeling results for robustness is a modus operandi of theoreticians.

These similarities clearly suggest that we should combine the expertise of philosophers of economics and philosophers of biology, who have independently puzzled over the nature of models and the epistemic value of robustness analysis. The debate amongst philosophers has mainly turned on whether robustness analysis can be a form of confirmation, with some arguing that it cannot be (Orzack and Sober 1993; Woodward 2006; Cartwright 1991) and others contending otherwise (Weisberg 2006; Kuorikoski et al. 2010). In this paper we combine the efforts of a philosopher of biology and a philosopher of economics to move this debate further. We will not argue that robustness analysis has no use. But we will argue that it is unable to confirm the sort of hypotheses that later figure in explanations of phenomena, several claims to the contrary notwithstanding. Robustness analysis should be viewed as a tool of discovery, not confirmation.

What sort of claims are at issue here? For robustness analysis to be a means of confirmation, we need to be clear what is purportedly being confirmed. Philosophers generally do not claim that it is models themselves. Rather, what can be confirmed are empirical claims of the form 'model *M* bears relation *R* to a target system *S*'. Philosophers have proposed different ways of articulating these claims and relations. Some speak of theoretical hypotheses and similarity (Giere 1988; Hausman 1992), others of isomorphism or partial isomorphism (Van Fraassen 1980; Da Costa and French 2003), others of capacities and isolations (Cartwright 1989; Maki 1992), etc. For the purposes of this paper, we merely assume that models are widely used to construct empirical hypotheses that later figure in explanations of particular biological or economic phenomena. Our focus here is the contribution of robustness analysis to the testing of these latter hypotheses.

Philosophers who defend the confirmatory power of robustness analysis do so because of the background knowledge that goes into the construction of models. We argue that even granting this input of background knowledge, still robustness analysis is not confirmatory per se. We give two reasons for this claim.

1. By and large only *some*, not all, of the idealizations of models are discharged by robustness analysis. As a result many robust theorems praised by theoreticians remain empirically questionable and thus explanatorily weak. We illustrate this fact with two examples: one from community ecology and another from auction theory. Crucially, this implies that these models cannot be read as specifications of causal mechanisms.
2. Robustness analysis crucially depends on showing that the assumptions of different models are independent of one another. However, we will argue that reports of their independence have been greatly exaggerated.

We then show that these two reasons make it implausible to claim that robustness results yield confirmation to causal explanations. If they don't do that, then what do they achieve? We argue that, at best, robustness results specify schemas that can then be used for building testable causal claims. We flesh out this idea using the notion of an open formula that one of us has defended elsewhere (Alexandrova 2008). Robust theorems can give us more fine-grained or elegant open formula or suggest interesting hypotheses to test, but no grounds for thinking that this open formula correctly specifies a given causal mechanism. For that we need to, first, fill in the open formula and then, second, to test it by conventional methods. So even when a given causal hypothesis is confirmed through empirical tests, the credit due the robustness of the model that inspired it, is far from clear.

We neither argue that robustness analysis is useless, nor that it should be abandoned. However, we find that its empirical value is easily exaggerated and overvalued.

In the section "Robustness analysis: a philosophical state of the art" we review the existing philosophical interpretations of robustness analysis in economics and biology. Section "Robust theorems in biology and economics" presents two examples of prized robustness results that nevertheless do not live up to philosophers' promises. Section "What robustness analysis does" presents our positive view.

Robustness analysis: a philosophical state of the art

The discussion of robustness analysis in philosophy of science starts with Richard Levin's classic statement in 1968 and William Wimsatt's subsequent elaboration in 1981. Levins identified the search for robust theorems as one of population biology's aims. He famously wrote, "Our truth is the intersection of independent lies" (1968, 423). Specifically, Levins claimed that a prediction of models which share a common set of biological assumptions but vary their idealizations, is robust. Wimsatt more generally identified robustness analysis with triangulation via independent means of determination, whereby different methods' convergence on the same result serves as the evidence of truth of this result (Wimsatt 1981). For example, one important example concerns the reliability of perceptual beliefs. We can triangulate the truth of a perceptual belief by using different sensory modalities. What is fundamental for both Levins and Wimsatt is that the different methods or models are independent.

Later, the empirical bona fides of robustness analysis were attacked by Orzack and Sober (1993) and Cartwright (1991). Orzack and Sober's critique consisted in identifying robustness analysis with an attempt to secure non-empirical confirmation, an enterprise that is inherently suspect. Suppose we have a family of models all of which entail some proposition. Why would this give us some reason to believe that proposition? Only if the set of models were mutually exclusive and exhaustive since then it would be a logical truth that the family of models imply the proposition. Otherwise, this is simply a theoretical examination of the consequences of the set. That is, the fact that a set of independent false models imply a proposition is not a reason to believe the proposition true. And as such, it is not a form of empirical confirmation.

Cartwright, on the other hand, questioned the analogy between robustness analysis and the well-accepted scientific methodology of detecting the same phenomenon with different methods. Consider Jean Perrin's attempts to measure Avogadro's number and to confirm the existence of atoms. Perrin used thirteen distinct methods to determine the number—by examining the Brownian motion of particles, Alpha decay, blackbody radiation, etc. It is the fact that all these methods yield approximately the same number that gives such a high confirmation to correct estimate of Avogadro's number and the existence of atoms. In this case a large number of independent, defeasible sources of evidence all point to the same conclusion. Wesley Salmon argued that this is an application of the common cause principle (Salmon 1984, 221). How do we know there are atoms? Because it is extremely unlikely that independent causes would yield measurements with such "remarkable agreement" with each other. Advocates of robustness analysis, Wimsatt for example, often use this argument too: the fact that a certain result is a robust theorem makes it likely that the false assumptions in these models do not matter and that the results are "driven" by the common assumptions. It is highly unlikely that the theorem is false given that the different models line up yielding these results.[1]

In this analogy justified? Not according to Cartwright. In the case of robustness, "different implementations [i.e., models] do not constitute independent instruments doing different things, but rather different ways of doing the same thing: instead of being unrelated, they are often *alternatives* to one another, sometimes even contradictory" (Cartwright 1991, 153, original italics). That a set of false models all imply the same result is indeed a coincidence, but not the sort of coincidence that would allow us to infer the truth of the result (ibid, 154).

Recently Weisberg 2006, Weisberg and Reisman (2008) and Kuorikoski et al. (2010) offered replies to the charges mounted by Orzack and Sober and Cartwright.

Weisberg responds to Orzack and Sober's critique as follows. Robustness analysis *by itself* does not deliver confirmation to explanatory hypotheses. Confirmation of said hypotheses occurs only when we can interpret the

[1] Robustness analysis is often used by scientific realists to argue for particular claims being true or certain entities as existing and bears a passing resemblance to the "no miracle argument". However, generally, robustness analysis is different than the no miracles argument. Robustness concerns the following conditional probability, $\Pr(O/M_1\&M_2\&\cdots\&M_n)$, and the no miracles argument concerns the following conditional probability, $\Pr(M/O_1\&O_2\&\cdots\&O_n)$. where O_i are predictions and M_i are models (or theories).

mathematical dependencies in the models as causal dependencies. Mathematical dependencies are not readily interpretable as causal because a theorem's robustness depends on assumptions biologists know to be false of real populations. However, we justify our belief that these assumptions do not matter in part through robustness analysis and in part through "low level confirmation—confirmation of the fact that certain mathematical structures can adequately represent properties of target phenomena" (Weisberg 2006, 851). A robust theorem gains low-level confirmation if it is successfully predicts the behavior of real systems and populations. This is why, the argument goes, the Volterra Principle has such a wide acceptance—it is a robust theorem *and* it has much empirical evidence to speak in its favor (Weisberg and Reismann 2008).

This response raises three issues. First, it is not clear by what criteria we determine whether a dependency is mathematical or causal. Second, the fact that a set of models have a common prediction does not entail that it is robust unless we have evidence that the assumptions, or at least a subset, of the models are independent and it is question-begging to suggest that the commonality of predictions provides this. Third, there is no reason to believe a proposition just because a set of false beliefs imply some proposition and thus mutatis mutandis with models' idealizations and predictions.

Recently, Kuorisoki, Lehtinen and Marchionni attempted to answer these worries. They argue that robustness analysis, even absent predictive successes, is an empirical method of confirmation. Kuorikoski et al. distinguish between three kinds of assumptions typically made by modelers: substantial, Galilean and tractability assumptions (Kuorikoski et al. 2010, 7).[2] The first state the putative causes of the phenomena of interest: for example, what species are predators or the nature of information available to an agent. The second kind idealize away putatively relevant factors not included in our model: for example, no transaction costs or no variation in mortality rates. The third kind are analytically useful ones: for example, the functions are twice differentiable, the distribution of prey is uniform, the equations have stable equilibria, and so on.[3] All three kinds of assumption are often involved in robustness analysis, but background knowledge of the empirical world clearly informs the substantial assumptions. For example, it is only because we already have evidence that the information available to an agent is relevant for his behavior that we include this assumption in the model. Or, we only include predation as opposed to competition because we antecedently have evidence that the prey species has no competitors. Clearly, at least sometimes, these beliefs are reasonable, true, and greatly informed by observation. Robustness analysis supposedly then explores under what conditions these putative causes have the effects that they do. This is some justification for interpreting dependencies in models as causal. But note that this does not amount to a specification of a causal mechanism, (for we still know that assumptions of our model are false and hence that the mechanism cannot be

[2] See also Musgrave (1981), Hindriks (2006), Cartwright (2009), Alexandrova (2006) and Maki (1992) for similar distinctions.

[3] It is not clear that these types of assumptions are that different. For example, a causal assumption that some factor C causes E assumes certain intervening factors are absent. But we leave worries like these to the side.

what the model says it is), let alone that this mechanism operates in the system of interest (more on this shortly). Still to the extent that this background causal knowledge enters into model building and into robustness analysis, we agree with Kuorikoski et al. that it is indeed not a non-empirical form of confirmation (11). But this is because we suggest robustness analysis not a method of confirmation at all. In particular, it does not amount to admitting robustness analysis can take us *beyond* our background knowledge. This takes us to the issue of independence and Cartwright's critique.

In responding to Cartwright, Kuorikoski et al. argue that nevertheless models that participate in robustness analysis are independent in the relevant sense. This relevant sense is the following: errors in one method of determination, i.e., in one model, are independent of errors in the other. For example, one tractability assumption may introduce a false proposition, another a different false proposition. But the likelihood (here defined as the subjective prior probability) that a certain result can be derived using one kind of tractability assumption, is independent of the likelihood that it can be derived using another. Thus the errors that different tractability assumptions might introduce are independent of each other as required by robustness analysis (Kuorikoski et al. 2010, 22).

The crucial thing to note about this defense is that it gives us no guarantee of independence of models: even if the same substantial assumptions conjoined with different tractability assumptions confer different likelihoods on a given prediction, this entails nothing about the independence of the tractability assumptions.[4] But does this differential likelihood make independence a reasonable bet nevertheless? We think not for the following reasons.

First, there is an important difference between Perrin's use of different empirical hypotheses all implying the same conclusion and a modeler's use of models each of which makes different assumptions that this modeler knows to be false. In the first case, we have several empirical hypotheses each of which has independent evidential support. In the second, we have several overlapping falsehoods. The fact that the modeler believes that the errors introduced by false assumptions are probabilistically independent, does not change the fact that these models are an essentially different beast from the various methods of determination used in classic forms of triangulation such as that discussed by Salmon. Consider a conjunction of a statement we believe to be false and one that we believe true. The probability of this conjunction is zero even when the statements are probabilistically independent since the subjective probability of a false claim is zero. Thus, if all or some of the Galilean and tractability assumptions are false, then replacing them with other false assumptions doesn't help. We shall be returning to this argument shortly.

Second, even bracketing the above problems and granting the relevant similarity between the case of Perrin and that of the models, still we might wonder if the subjective probability that the errors are independent is sufficient for the sort of independence of evidence that triangulation requires. When carrying out robustness

[4] Suppose $\Pr(O/C\&A_1) \neq \Pr(O/C\&A_2)$ where O is some prediction, C is the common "substantial" core, and the A_is are different idealized assumptions. It follows that $[\Pr(O)\Pr(C\&A_1/O)]/\Pr(C\&A_1) \neq [\Pr(O)\Pr(C\&A_2/O)]/\Pr(C\&A_2)$. Thus, $\Pr(C\&A_1) \neq \Pr(C\&A_2)$. However, this is consistent with $\Pr(A_1/A_2) \neq \Pr(A_1)$.

analysis, modelers do not randomly pick different Galilean idealizations and tractability assumptions from a large sample of all such assumptions. Rather, they select whatever assumptions work with their methods, abilities, training, etc. The fact that a prediction is common over a suite of models can sometimes be due to the fact that the models were designed to make just that prediction. For example, it can be the case that a prediction is common because of the mathematical framework used. Even if they think the errors introduced by each of these assumptions to be independent of each other (which we doubt), this belief is very likely to be wrong. Try deriving your robust result without any tractability assumptions at all—it simply won't work. So how can the fact that your standard toolbox yielded a couple of different tractability assumptions that do imply the result provide genuinely independent evidence?

For these reasons we are skeptical of the "independence argument". But our main reason to doubt the confirmatory power of robustness analysis is that there is no reason to think that robust theorems by themselves provide adequate representations of actual causal relations. Weisberg uses the term "adequate representation" to refer to a robust theorem with some degree of "low-level confirmation". This may sound like a claim that adequate representations state causal relationships, but we insist that they needn't. Even when a theorem has been shown to be robust to some idealizations, we may not have a causal relationship represented. When articulating a causal claim we need to specify the conditions in which the causal relation holds. However, for each model, there are assumptions we know to be false or whose truth we cannot evaluate (i.e., tractability assumptions). Hence, we do not know what the actual conditions are which, when conjoined with the causal factor of interest, produces the predicted effect. Put differently, unless we can "de-idealize" our Galilean assumptions (and others if we cannot sharply distinguish between different types of idealized assumptions), we do not know that we have adequately represented a causal relationship.

Of course, the advocates of robustness analysis may claim if a robust theorem yields a conditional claim, then this conditional claim specifies the causal mechanism responsible for the phenomenon described by the robust theorem. The antecedents represent putative causes and enabling conditions and the consequents putative effects. For example, the Volterra Principle is a representation of a causal mechanism: "Ceteris paribus, if the abundance of predators is controlled mostly by the growth rate of the prey and the abundance of the prey controlled mostly by the death rate of predators, then a general pesticide will increase the abundance of the prey and decrease the abundance of predators" (Weisberg 2006, p. 737).[5] Kuorikoski et al. similarly argue the common structure of a family of models is a representation of a causal mechanism that explains the relevant phenomenon (2010, 14).

But in order to defend this interpretation of robust theorems, the advocate needs to claim that all the false assumptions of the family of models in question have been

[5] Our views should not be construed as skepticism regarding causal claims. For example, we have reason to believe the Volterra Principle is a true causal claim when we can intervene in a predator–prey system and produce the intended effect. For example, all things being equal, if we instantiate the antecedent and the consequent is observed that is reason to believe the Volterra Principle describes a causal relationship. However, if we do this, robustness analysis plays no role in confirmation.

discharged. Rarely can this be done. To see why this so, let us introduce some helpful terminology. Consider a set of distinct models $\mathbf{M} = \{M_1, M_2, \ldots, M_n\}$ where each model is composed of a common substantial "core" and distinct auxiliary assumptions of the Galilean or tractability sort. Thus, we can construe each model as $M_i = (C\&A_i)$. Let us that say that M_i and M_j with C are distinct if, and only if, they contain logically non-equivalent auxiliaries A_i and A_j (where $i \neq j$). Finally, a prediction P is robust over $\mathbf{M}$ if, and only if, for each M_i in $\mathbf{M}$, M_i implies P; otherwise, P is fragile. For simplicity we are considering models as sets of propositions and the relation between models and predictions as deductive, but nothing hangs on this. Let's suppose for any P and idealized A_i such that P follows from $(C\&A_i)$, we can provide another A_j such that P follows from $(C\&A_j)$. It is the case that either A_j is idealized or it is not. If it is idealized, then we must find some other assumption to discharge P with respect to A_j and so on unless there is at least one of auxiliary assumptions, which is true. Otherwise, though we discharge individual idealized auxiliaries we cannot discharge the set of idealized auxiliaries. Thus, to discharge our skepticisms about A_i our robustness analysis must terminate in a set of auxiliaries, which must have at least one true member.[6] Given the ineliminability of idealizations, robustness analysis cannot quiet worries about them.

Analogously, suppose one is worried about the truth of some perceptual belief formed based on one sensory modality. We can confirm the truth of the belief by consulting another sensory modality given its independence and reliability. However, if one is worried about the reliability of sense perception per se, then there is no other sense modality to consult. Moreover, in the case of idealizations, we know that the alternative assumptions are false.

Another way of seeing the issue is to compare *relative* versus *absolute* robustness analyses. Suppose we are worried about an idealized assumption and that our prediction depends on it along with the substantial core. We can remove this worry by replacing it with another assumption, which in conjunction with the substantial core implies this prediction. Call this the "relative" robustness analysis. Suppose we are worried about *all* of the idealized assumptions that when conjoined with the common core imply our prediction. The only way to remove this worry is to show that there is some true assumption when conjoined with the substantial core that implies the prediction. Call this the "absolute" robustness analysis. Our worry reduces to this: if we are worried about idealizations simpliciter we need absolute robustness analyses and when we are worried only about certain idealizations relative robustness analyses are sufficient. However, robustness analyses only have confirmatory value when we can perform absolute robustness analyses.

We can now embed this discussion with regard to our concerns about causal hypotheses. Recall the distinction between Galilean idealizations and tractability assumptions. The former idealize away disturbing causal factors, the latter are mathematical niceties that allow derivation (elsewhere one of us referred to these as derivation-facilitators—Alexandrova 2006). To evaluate a conclusion's robustness

[6] A common response to this sort of argument is that we employ a notion of approximate truth and suggest that say A_j is closer to the truth than A_i. However, as far as we are aware, no such account of approximate truth is successful. And in any case, many of the false assumptions in question are not even close to truth on any understanding of approximate truth.

with respect to Galilean idealizations, we might replace these assumptions with realistic ones. This procedure called 'de-idealization' was famously described by McMullin (1985). For example, if a model assumes no transaction costs, we take these costs taken into account. Likewise, in predator–prey models we can replace the linear functional response with more realistic functional forms. So, the first reason that most principles called robust theorems in economics and biology cannot be read as specifying actual causal mechanisms is because *not all* of the Galilean idealizations can be discharged in this way. In economics, we cannot discharge the assumption of utility maximization across many contexts, to use one example. Often de-idealization introduces so much complexity that models do not have analytic solutions or don't predict a particular outcome. If de-idealization is not available for some of the Galilean idealizations, then we have no grounds for claiming that the falsity of these assumptions does not matter, and so no ground, beyond that of our background knowledge that informed the model, for claiming that the model specifies a causal relation.

Kuorikoski et al., aware of this problem, qualify their interpretation as follows: robustness analysis may not be able to find mechanisms, nor establish them conclusively, but it gives us *some* confirmation that the common structure of a family of models identifies a genuine causal mechanism (Kuorikoski et al. 2010, 17, 20). Some support is still support.

We reject the claim that as different idealizations are discharged, the model's hypothesis acquires more and more credibility. That would be the case if a) models were independent and b) we had reasons to overlook the falsity of their assumptions.

In the absence of this, our question to Kuorikoski et al. is as follows: Some support for what? If we know that some parts of our causal mechanism specification are false, then it is unclear what sort of claim we are trying to confirm by investigating the robustness of the model which purports to specify this mechanism. We agree that robustness analysis can provide support for the claim that a proposition implied by a model doesn't depend on some false propositions. We also readily acknowledge the role of background knowledge in model construction. We even grant their claim that robustness analysis helps us to differentiate the relative importance of various assumptions within a model (Kuorikoski et al. 2010, 18). However, none of this implies that the robust theorem describes a causal mechanism. To confer that, a robust model needs to tell us what the mechanism is without surreptitiously ignoring all the idealizing and tractability assumptions that have not been discharged. This is what we, along with others, don't see happening.[7] Relative robustness analysis is able to show that some idealizations do not matter, but by no means all. Thus, we remain skeptical that robustness analysis can contribute to the confirmation of causal claims.

Before we offer our own account of what robustness analysis provides, we present two striking examples of robust modeling results that display our worries. These examples and their like are not unusual or atypical and their role deserves to be studied.

[7] Julian Reiss has recently made this point forcefully for economics (Reiss 2008 118–119).

Robust theorems in biology and economics

An example from biology

Succession is the compositional and structural changes that occur in communities and ecosystems through time. Ecologists say that succession is *primary* when it involves the colonization of an area absent an existing ecological community and it is *secondary* when it involves an already existing community after some disturbance. With regard to forest secondary succession, ecologists have noted that similar initial communities proceed with similar successional stages, different initial communities converge on similar final states, and these final states resemble undisturbed forests. Ecologist Henry Horn has studied forest succession by modeling them as "regular Markov chains" (Horn 1975).

A finite Markov chain is a process with a finite number of states in which the probability of being in a particular state at step $n + 1$ depends only the state at step n and it is regular when in addition if for some power of the transition matrix it has only positive entries. To model forests in this way, Horn assumes that the probability that a given species will be replaced by a species is proportional to the number of saplings of the latter in the understory of the former. Thus, the abundance of a species at $t + 1$, is the sum of the species abundances at t multiplied by the probability that they are replaced by the focal species. As the number of generations gets larger and larger, the community will arrive at a stationary distribution where species replace each other with characteristic frequencies.

Horn himself recognizes that his model is highly idealized and engages in a robustness analysis when he writes,

> I shall routinely make outrageous assumptions, but I shall defend them in several ways. Some are needed only for analytic convenience and may be relaxed with no major effect on the result. In some cases a redefinition of a measurement is all that is needed to bring theory into line with fact. Astoundingly, some of the assumptions are even true (1975, 197).

He makes a large number of idealized assumptions including communities can be represented with discrete states, the transition matrix is homogenous (the transition probabilities are constant), there is no spatial structure, there is no density-dependence, there are a large number of patches, no time lags, and trees die and are replaced at the very same time. Using his model and data collected from forests around the Institute for Advanced Studies at Princeton University, Horn demonstrated that the vector of observed abundances and that predicted by the model were in close agreement. This was an impressive result for such a simple model.

When evaluating models, it is common to test them in several ways (Lloyd 1994). First, we consider the model's fit to empirical data. Second, we determine if the assumptions of the model are true of the system of interest. Third, we evaluate both fit and the assumptions' truth across a variety of systems. Horn had provided evidence for fit but what about the second criterion? Given that he antecedently knew many assumptions are false, no confirmed predictions would confirm them.[8] Horn recognized the assumption that trees die and are replaced at the very same time is false and he replaces that assumption with one where life spans vary among tree species and change continuously through time. Moreover, he shows that the results are robust with regard to this change in assumptions. Unfortunately, though he removed doubts regarding one assumption, we have seen that there is a whole host of other idealizations on which no robustness analysis is carried out. If this is so, then the following analogy is weak,

> The most dramatic property of succession is its repeatable convergence on the same climax community from any of many different starting points. The property is shared by a class of statistical processes known as 'regular Markov chains' (Horn 1975, 196).

Example from economics

Auction theory models strategic interaction of rational agents in various kinds of auctions. Auctions differ among each other by their rules. These affect the information available to the bidders and these bidders' incentives. For example, an auction might be first-price or second-price. In the former the winner pays the price equivalent to the highest bid, while in the latter the second-highest. How do these auctions compare with respect to the revenue generated? A set of central results in auction theory—Revenue Equivalence Theorems—show that under particular conditions these two kinds of auctions are equivalent with respect to the profit they generate for the bidders and the revenue they bring for the seller. Here is its statement:

> Assume each of a given number of risk-neutral potential buyers of an object has a privately-known signal independently drawn from a common strictly increasing, atomless distribution. Then any auction mechanism in which (i) the object always goes to the buyer with the highest signal, and (ii) any bidder with the lowest feasible signal expects zero surplus, yields the same expected revenue (and results in each bidder making the same expected payment as a function of her signal) (Klemperer 1999, 232).

Many auction models in equilibrium, award the object to the bidder with the highest value, i.e., a bidder's probability of winning is just the probability that her valuation is the highest. Similarly, in all these models the bidder with the lowest valuation has no chance to profit, so, the theorem says, all these mechanisms will yield the same economic result. The result applies to common as well as private values (that is, when the value of the good for sale is known to all and when it is known only to the bidder herself), and can also be extended to cases in which more than one object is for sale.

However, revenue equivalence cannot be proved once we leave the simple world of independent draws from strictly increasing atomless distributions. Independence of signals and specific properties of distributions are necessary assumptions for the

[8] That is, given his prior probability with regard to those idealizations was zero, then regardless of the likelihoods, the posterior probability must be zero.

revenue equivalence results. It means that if players receive information from non-independent sources—that is their valuations are affiliated—revenue equivalence does not follow. This is one reason why the theorem has *not* been influential in empirical work on auctions. Klemperer, an eminent auction theorist and real-world auction designer (he advised the UK government on their 3G auctions in 2000), makes a striking observation: for all the theoretical sophistication of the extensive and technical auction literature, it is of little use to practical auction design (Klemperer 2002). Revenue Equivalence theorem is a celebrated result and a center piece of auction theory. If there are robust theorems in economics, surely this is one of them. But due to its reliance on assumptions that cannot be discharged, the Revenue Equivalence Theorem fails to specify a causal mechanism.

We contend that this is not an atypical situation. Even the example that Kuorikoski and his co-authors tout to be a success case for robustness analysis—the Center-Periphery models in Geographical Economics—fails to qualify as a specification of a causal mechanism, simply because, by their own admission, only some of the Galilean assumptions get discharged (Kuorikoski et al. 2010, 17, fn 20).

What robustness analysis does

If robustness analysis on its own does not specify nor confirm causal relationships, then how should we "read" robustness theorems? And what exactly does robustness analysis achieve? These are our questions for this section. We start with the first one.

The idea that models serve as templates or incomplete schemas for empirical claims has been around for some time. Perhaps the first explicit connection was made by Sidney Morgenbesser in his 1956 PhD dissertation at the University of Pennsylvania. He referred to claims of social science as "schematic expressions" that do not commit social scientists even to singular explanations, let alone universal generalizations (Morgenbesser 1956, ch. 2). Rather, schematic expressions are frameworks that can, with additional work, provide genuine explanations. Later Paul Humphreys and William Wimsatt each independently used the term 'template' to describe models across the mathematical sciences. For Wimsatt, neutral models are templates for further important explanatory and exploratory work in virtue of omitting a variable or assuming its value is zero (Wimsatt 1987). For example, in neutral models in evolutionary theory and ecology, we assume respectively that traits do not differ in their fitnesses or there is no competition between species. For Humphreys, a template is a syntactic structure that expresses some relations between some variables without interpreting them or assigning specific values (Humphreys 2002).[9] Many philosophers of economics and biology express the idea that models, idealized or with unrealizable assumptions, are nevertheless *useful* for building good explanations (Odenbaugh 2005; Hamminga 1983 among many others). Most recently, philosophers have suggested that models in economics and

[9] There is also a tradition in the semantic view of theories for regarding models as providing theoretical "definitions" or set-theoretical predicates. Thus, syntactically or semantically, we can separate the model from the model's application.

biology propose *possible* or *potential* explanations (Gruene-Yanoff 2009; Aydinonat 2008; Reiss 2008; Forber forthcoming; see also Brandon 1990). They do not, at least generally, supply actual explanations, however.

There are many ways to make good on these related ideas, and here we employ one of them: robustness theorems provide us with open formulae that can be used to build hypotheses about mechanisms, and robustness analysis is a way of chiseling out these open formulae.

Interpreting models as open formulae has been defended recently by one of us (Alexandrova 2008). But in this paper, we do not assume its correctness and use it merely as a vehicle for specifying the contribution of models, given our view that models should not be read as statements of causal mechanisms even after robustness analyses.[10] Suppose we treat a model as giving us (among other things) a rough template for a causal claim, but not an actual causal claim of the form "feature(s) F(s) cause behavior(s) B(s) under condition(s) C(s)". Such a rough template might look like this:

(OF) In a situation x with some characteristics that may or may not include $\{C_1, \ldots, C_n\}$, a certain feature F causes a certain behavior B.

We might call this template an open formula because a) it does not quantify over x and b) it does not specify x (whether or not it fully specifies 'F', 'C_i' and 'B'). That is, there is neither commitment to x existing, nor yet any claim about any phenomenon since the features of x are not specified. x is a free variable, which needs to be filled in in order for the open formula to make such a claim, or to specify a proposition. Once x is specified, we get a proper causal hypothesis (and of course when 'F', 'C_i', and 'B' are as well).

The point of reading models in such a weak way is to accommodate the "facts of life". Given the lack of de-idealization we have no grounds for thinking that models even after a robustness analysis give us statements of causal mechanisms, nor for thinking that models can tell us even a single full set of conditions under which a given causal relation suggested by a model holds. Although a model may not tell us these things, it could still supply the categories Fs, Bs and some Cs, which we can use to construct causal hypotheses. This is indeed how good application of models proceeds—applied economists, for example, design institutions and explain their functioning using partly models, partly experiments and much more.[11] Likewise, it is in experimental biology that we form and evaluate causal hypotheses through interventions on variables of interest.[12]

[10] This implies that we also reject the view that models state capacity claims (Cartwright 1999) or make claims about credible worlds (Sugden 2000, 2009). For reasons to reject these views see Cartwright (2009), Gruene-Yanoff (2009), Alexandrova (2008), Alexandrova and Northcott (2009), and Odenbaugh (2006).

[11] See Al Roth, the father of design economics (among many others, of course) on the complex mixture of methods that goes into successful applied economics (Roth 2002).

[12] "Natural experiments" are important in both economics and biology. Here one finds to situations such that some purported causal factor is present in the former and not the latter and where everything else is roughly the same. The problem with these so-called experiments is that rarely are they even roughly the same. Still, they can be important.

When a modeling result is a robust one, the open formula that corresponds to it is formally the same as an open formula of a regular model: it still has at least one free variable and still fails to fully specify even a single set of conditions under the causal relationship between its *F*s and *B*s holds. Is it nevertheless a better open formula than one of a non-robust model? On the negative side, an open formula of a robust set of models is no more likely to underwrite a causal hypothesis that turns out to be true than an open formula of a non-robust model (holding fixed the evidence supplied by the relevant background knowledge). However, an open formula arrived at via robustness analysis may have the following virtues:

1. *Independence of particular idealizations*. We grant that robustness analysis can resolve worries about specific unrealistic assumptions by discharging these assumptions when there are other independent assumptions that do the same work. But because we can rarely discharge them all, we rarely get out of the "circle of idealizations" and hence we may not claim that a robust modeling result is independent of all idealizations. Thus, we cannot specify exactly what the causal relation between *F*s and *B*s and discharging idealizations does not amount to confirmation.
2. *Greater generality*. Likewise, as we remove irrelevant conditions we can arrive at more and more general open formula. For example, we may be able to eliminate some of the C_is which we originally believed were causally relevant in the relationship between *F*s and *B*s. If this is so, our open formula becomes more and more austere or if you like "simple".

Conclusion

Robustness analysis can be an excellent formal technique for identifying open formulae. As such, it is a good tool of discovery. It takes certain bits of our background knowledge, disciplines them with a formal treatment, and can even tell us that these bits do not depend on particular falsities. However, absent more empirical input, it does not take us any further. We believe that its virtues as tools of discovery have been misinterpreted as virtues of confirmation of empirical claims.

Acknowledgments Authors would like to thank Michael Weisberg, Robert Northcott, Nancy Cartwright, Uskali Maki and two anonymous referees for useful comments on the paper.

References

Alexandrova A (2006) Connecting rational choice models to the real world. Phil Soc Sci 36(2):173–192

Alexandrova A (2008) Making models count. Phil Sci 75:383–404

Alexandrova A, Northcott R (2009) Progress in economics: lessons from the Spectrum Auctions. In: Kincaid H, Ross D (eds) Oxford handbook of philosophy of economics. Oxford University Press, Oxford, pp 306–336

Aydinonat NE (2008) The invisible hand in economics: how economists explain unintended social consequences. Routledge, London

Brandon R (1990) Adaptation and environment. Princeton University Press, Princeton

Cartwright N (1989) Nature's capacities and their measurement. Oxford University Press, Oxford
Cartwright N (1991) Replicability, reproducibility, and robustness—comments on Harry Collins. Hist Polit Econ 23:143–155
Cartwright N (1999) The dappled world. Cambridge University Press, Cambridge
Cartwright N (2009) If no capacities, then no credible worlds: but can models reveal capacities? Erkenntnis 70:45–58
Da Costa N, French S (2003) Science and partial truth: a unitary approach to models and reasoning in science. Oxford University Press, New York
Forber P (forthcoming) Confirmation and explaining how possible. Stud Hist Phil Sci C Stud Hist Phil Biol Biomed Sci
Giere RN (1988) Explaining science. University of Chicago Press, Chicago
Gruene-Yanoff Till (2009) Learning from minimal economic models. Erkenntnis 70:81–99
Hamminga B (1983) Neoclassical theory structure and theory development. Springer, Berlin
Hausman DM (1992) The Inexact and separate science of economics. Cambridge University Press, Cambridge
Hindriks FA (2006) Tractability assumptions and the Musgrave–Mäki typology. J Econ Methodol 13:401–423
Horn H (1975) Markovian properties of forest succession. In: Cody ML, Diamond J (eds) Ecology and evolution of communities). Belknap Press, Cambridge, pp 196–211
Humphreys P (2002) Computational models. *Philosophy of Science* Supplement: Proceedings of the 2000 Biennial Meeting of the PSA. Part II: Symposia Papers, 69/3: S1–S11
Klemperer P (1999) Auction theory: a guide to the literature. J Econ Surv 13(3):227–286
Klemperer P (2002) What really matters in auction design. J Econ Lit 16(1):169–189
Kuorikoski J, Lehtinen A, Marchionni C (2010) Economic modelling as robustness analysis. Brit J Phil Sci 23:5–7. doi:10.1093/bjps/axp049
Levins R (1968) Evolution in changing environments: some theoretical explorations. Princeton University Press, Princeton
Lloyd E (1994) The structure and confirmation of evolutionary theory. Princeton University Press, Princeton
Maki U (1992) "On the method of isolation in economics. Poznan Stud Phil Sci Humanit 26:319–354
Maki U (1998) Aspects of realism about economics. Theoria 13:301–319
McMullin E (1985) Galilean idealization. Stud Hist Philos Sci 16(3):247–273
Morgenbesser S (1956) Theories and schematas in social sciences. PhD Dissertation, University of Pennsylvania
Musgrave A (1981) Unreal assumptions in economic theory: the F-twist untwisted. Kyklos 34:377–387
Odenbaugh J (2005) Idealized, inaccurate, and successful: a pragmatic approach to evaluating models in theoretical ecology. Biol Philos 20:231–255
Odenbaugh J (2006) Struggling with the sciences of ecology. Biol Philos 21:395–409
Orzack SH, Sober E (1993) A critical assessment of Levins's The Strategy of Model Building in Population Biology (1966). Q Rev Biol 68:533–546
Reiss J (2008) Error in economics: the methodology of evidence based economics. Routledge, London
Roth A (2002) The economist as engineer: game theory, experimental economics and computation as tools of design economics. Econometrica 70(4):1341–1378
Salmon W (1984) Scientific explanation and the causal structure of the world. Princeton University Press, Princeton
Sugden R (2000) Credible worlds: the status of theoretical models in economics. J Econ Methodol 7(1):1–31
Sugden R (2009) Credible worlds, capacities and mechanisms. Erkenntnis 70:3–27
Van Fraassen BC (1980) The scientific image. Oxford University Press, Oxford
Weisberg M (2006) Robustness analysis. Phil Sci 73:730–742
Weisberg M, Reisman K (2008) The robust Volterra principle. Phil Sci 75:106–131
Wimsatt WC (1981) Robustness, reliability and overdetermination. In: Brewer MB, Collins BE (eds) Scientific inquiry and the social sciences. Jossey-Bass, San Francisco, pp 124–163
Wimsatt W (1987) False models as a means to truer theories. In: Nitecki M, Hoffman A (eds) Neutral models in biology. Oxford University Press, London, pp 23–55
Woodward J (2006) Some varieties of robustness. J Econ Methodol 13(2):219–240