Robustness and Replication: Models, Experiments, and Confirmation

Gareth Fuller
June 19th, 2022

This is a draft. Comments are welcome.

Email: gfuller2@ku.edu
Orcid ID: **0000-0002-0879-8198**

Abstract: In this paper I take up a criticism of robustness analysis. Robustness analysis is a method of confirming idealized models where several models with a shared core causal mechanism but different idealizations are constructed. If this set of models produce a shared result, then this is supposed to show that the core mechanism is what is responsible for this response. If this shared result is empirically confirmed, it is argued that this then provides confirmation for the shared core. Robustness analysis has faced several criticisms, one being that, since each model in the robust set is false, no confirmation can be had. Even if they all agree, each model is still false, and agreement among false models confirms nothing. I argue that this concern can be assuaged by understanding the role fo some idealizations as controls for causal factors. I draw an analogy between such idealizations and contrived experimental conditions. From there, I extend the analogy so that robustness analysis can play a similar role to some cases of experimental replication. I finish by considering some concerns about this analogy and some residual fears about idealizations.

Introduction

The role of robustness analysis of models is a significant practice in many sciences, such as ecology, climate science, and economics. Its role is particularly significant where other, more directly empirical means of confirmation may be unfeasible or impossible. However, the confirmatory power of robustness analysis has been readily questioned in a variety of ways.[1] In this paper I attempt to resolve a dilemma for the confirmatory power of robustness analysis initially presented in Orzack and Sober and expanded upon by Odenbaugh and Alexandrova.[2] The dilemma holds that robustness analysis cannot be

[1] Orzack and Sober 1993, Sugden 2001, and Odenbaugh and Alexandrova 2011
[2] Orzack and Sober 1993, Odenbaugh and Alexandrova 2011.

confirmatory since it deals with false, idealized, models and false models cannot be confirmed. However, if a model is de-idealized robustness analysis is no longer needed. I agree with these critics that many claims of the confirmatory power of robustness analysis are overstated, but that there are conditions whereby robustness does provide confirmation of a modeled result. I defend this claim by drawing an analogy between robustness in modeling and replication in experiments.

In setting out my account of robustness analysis, and developing the analogy to experiments, there will be some alterations to the structure of robustness analysis as it was originally laid out in some early texts (Levins 1966, 1968). In particular, I will argue that conclusions from robust analyses will often need to be more constrained than they are. While this might make my defense of robustness analysis less interesting, I will sketch a way that the stronger conclusions that are often the goal of robustness analysis might be built out of multiple robustness analyses.

The paper will proceed as follows. In section 1 I present the basics of robustness analysis and the proposed dilemma. In section 2 I consider the dilemma in relation to replication in experiments. In section 3 I draw an analogy between replicating experiments and robustness in models. In section 4 I consider some concerns over this analogy. In section 5 I present the concessions and upshot of my position.

1. The Background of Robustness

Much of the philosophical debate on robustness analysis was spurred on by a paper written by biologist Richard Levins (1966). Levins' paper covered several related topics on the construction of models in population biology. While discussing various trade-offs that biologists need to make to build models, Levins comes to the conclusion that any particular model will be missing some key aspect of the target that will make it a poor representation of its target. Levins goes so far as to refer to individual models as "lies".

Models need to make trade-offs and are lies for a variety of reasons. As Levins points out, a key factor is the finitude of human cognitive power (even when aided with external cognizers like computers). Any model in population biology that were to accurately capture *all* of the causally relevant factors for

some phenomenon would be many differential equations, impractical to deal with, unlikely to be understandable to humans, and mathematically intractable. This is no way to go about studying the world, so some trade-offs are necessary.

Now, in many cases we might be forced to use models to study the world because we are limited in other respects. In the case of population biology, as well as climate science and economics, we might not have much other epistemic recourse. The phenomena are too big, temporally extended, or otherwise inaccessible to be amenable to other epistemic approaches. Further, models have found their way into the practices of many sciences for a variety of reasons, and so there is little reason to push them out.

Given that we will use models, and that they must be simplified to be of any use to humans, models are built with the goal in mind to accurately capture only some aspects of their target. Weight consider optimality models in biology. These models may be used to provide an explanation for the prevalence of some phenotype in a given population. However, optimality models do not include all of the causally relevant factors for the existence of some phenotype, as they do not have some way of representing the genotypes of a population. Optimality models, however, are able to represent selective pressures that might lead to

It is in this way that models are lies. Although they capture some aspects of the world correctly, they simplify, abstract, and remove much of what is going on in the target system. The real concern, for our purposes, are the kinds of simplification that count as idealizations. An idealization is the intentional remove of some causally relevant factor. In the case of LV, this may include information about the density relationship between predator and prey, carrying capacity of an environments, and many other factors that are not always represented.

To resolve this problem, robustness analysis was introduced.

The starting point for robustness analysis is that models used in sciences incorporate many falsehoods. Models are generally considered to be made up of sets of assumptions, some of which are meant to be accurate, other assumptions not so much.[3] Some of these less accurate assumptions are

[3] Schulz 2015.

generally viewed as rather harmless.[4] However, some assumptions remove or falsify causally relevant factors in the target system and are viewed with greater suspicion. These will be referred to as idealizations and will be the focus of the discussion below.

Idealizations are recognized as a common and integral part of models. Systems that are modeled by scientists are often deeply complex and incorporate many different causal influences. As Levins points out, modeling a population in complete detail would require capturing "genetic, physiological, and age heterogeneity within species of multi-species systems that are changing demographically and evolving under fluctuating influences from the environment."[5] To create such a model would require hundreds of differential equations and would be intractable.

That models are idealized is not problematic on its own, but given this complexity it often not clear whether an idealization impacts a model.[6] As well, in many fields that employ robustness analysis, there is often limited theoretical oversight to lend a hand. Ecology, for instance, lacks the overarching laws that might be found in Newtonian Mechanics.[7] Given this lack of theoretical oversight, it is often hard to distinguish *how much* influence a particular idealization might have.[8] For instance, given some calculation on a frictionless plane, it is possible to calculate how the result might change by adding friction into the model. However, given an optimality model for some behavioral phenotype, it is not possible to add genetic influences into such a model.[9]

So, given the necessity of idealizations and the general complexity of many models, it is not clear whether some result derived from the model is a result of the assumptions of the model that are meant to

---

4 What counts as inaccurate but innocent assumptions is not agreed upon, but it is generally agreed that there are some such assumptions. Even those who argue against idealized models as being confirmatory do not hold that *all* false assumptions are problematic (Alexandrova 2008, Odenbaugh and Alexandrova 2011).

5 Levins 1966 pg. 421

6 Levins 1966

7 Levins 1966

8 Kurikoski *et al*. discuss some reasons why this is the case in economics, particularly when compared to other sciences like physics. Given the nature of the parameters, the fact that they are constantly varied for instance, there are no ways to perform an analysis of how much a certain value of the parameter throws off the results, outside of the proposed robustness analysis.

9 Levins 1966, Potochnik 2010. This is for several reasons, one of which is that optimality models do not have a parameter set aside for genetic influences. Optimality models are designed to capture how selection pressures effect phenotypes, and do this without considering genotypic influences.

be accurate or if they are driven by some idealization. The concern is that, if we want our models to tell us about the world, then we want the results to be derived from the parts of our models that are like the world. The strategy of robustness analysis is to develop several models with a shared core set of assumptions, but different idealizations. If all of the models produce the same result, then this is supposed to lead to the conclusion that it is the shared core of the models that is responsible and not the idealizations. Such a result shared by all the models in the robust set is known as a "robust theorem".

Take some idealized model specifying an unconfirmed causal mechanism that implies some empirically confirmed result. Given the assumed complexity of the model, it is unclear if the result was produced by the, hopefully, accurate assumptions stipulating the core mechanism, or if it was driven in some key way by the idealizations. Since the idealizations are known to be false, this means that, even if the model produces the empirically confirmed result, this does not confirm that the model accurately represents the causal behavior of the target system. The idealizations remove some causally relevant factors, after all. Removing the idealizations may not be feasible or possible, so the next best option is to build a model with a different set of idealizations. If the empirically confirmed result turns out to be a robust theorem, this is meant to show that the result is driven by the shared core and therefore may provide evidence for the mechanism specified therein.

We might consider the example of robustness analysis from Levins (1966). In this example, Levins considers three models, and takes as a robust theorem that, "in an uncertain environment, species will evolve broad niches and tend toward polymorphism."[10] Two of the models that Levins considers are optimality models, while one is a genetic model. The two optimality models differ in several regards, one model treats different environments as discrete and the other as continuous. Further, the model with discrete environments holds that fitness in one environment decreases fitness in the others, while the model with continuous environments holds that fitness adaptations to one environment do not impact fitness in other environments. These are idealizations since it is not the case that all such adaptations to an environment will impact fitness to others. Further, both ignore genetic factors, being optimality models,

[10] Levins 1966, pg. 423.

and so the third model in Levins' robust set is a genetic models. Genetic models, however, ignore selection pressures that are captured by optimality models. So, each model in the robust set trades out some idealizations found in the others.

There have been several criticisms of robustness analysis. For example, some have raised questions about the possibility of the kind of non-empirical confirmation provided by robustness analysis, while others have argued that models are not appropriately epistemically independent of each other.[11] However, here, I focus on what is arguably a more central criticism—what I call the "confirmatory dilemma".

It is important to distinguish how robustness analysis has been understood to work from the (mere) accumulation of multiple bits of evidence. The two are similar in that they are held to increase support for some proposition. When it comes to accumulating evidence, everything else being equal, if we have two bits of evidence for some proposition, we have more reason to believe it.[12] One reason being that evidence is fallible. So, if we have two bits of evidence for some proposition, and it turns out that one of the bits of evidence is flawed, then we still have one bit of evidence remaining.[13]

Robustness is not meant to work this way. First, each model in the robust set is false, and so, if the purpose of any of the models is to confirm the causal core, no individual model can do this. It is the complete set of models, each showing that the idealizations of the others are irrelevant, that is intended to confirm the causal core. So, we start with several bits of non-evidence and by combining them together in confirmational alchemy we get evidence.

This sets up the confirmatory dilemma.[14] Since all of the models in a robust set are false, it can provide no confirmation. Since each model is idealized, none of them specify a causal mechanism, and

---

[11] See Orzack and Sober 1993 for a criticism of robustness as non-empirical. See Cartwright 1991, Orzack and Sober 1993, Odenbaugh and Alexandrova 2011, and Justus 2015 for concerns about independence of models.

[12] Achistein 2001.

[13] We might consider this in the case of two independent experiments that seem to confirm some hypothesis. If it turns out that one of the experiments was conducted using flawed procedures, we still have the other experiment as evidence for the hypothesis.

[14] Orzack and Sober 1993, Odenaugh and Alexandrova 2011.

certainly not a causal mechanism as it appears in the world.[15] So, even if the robust set shows that the shared core drives the results of the model, an actual causal mechanism still has not been modeled. In order for the robust set to provide confirmation, at least one of the models needs to be de-idealized. However, once a model is de-idealized there is no need to discharge the idealizations, and so no need for a robust set. Either way, robustness analysis provides no confirmation.

So, we end up with a dilemma for robustness analysis. On the first horn of the dilemma a robust set provides no confirmation because each model in the set is idealized. However, on the other horn, a de-idealized model undercuts the need for a robust set. Therefore, the argument goes, robustness has nothing to add in terms of confirmation.

2. Robustness and Experimental Replication

The role of robustness reasoning in scientific methods other than models has received fair philosophical attention as well.[16] In particular, discussions have included the use of multiple experiments. It is often held that multiple, independent, experiments be run to boost confirmation of some experimental result through replication or reproduction of the experiment.[17] When it comes to replicating an experiment it is not always strictly about exact reproduction as this is often an impossibility.[18] An attempt at replication is "the repetition of what is presumed to matter for obtaining the original result".[19] This can be carried out by trying to single out what is taken to be the causal mechanism responsible for the initial result and testing it under appropriate, but different, conditions. The reason that replication is important is because there is uncertainty about what drives the result of a single experiment when treated on its own.[20]

---

15 Odenbaugh and Alexandrova 2011, pg.764.

16 Wimsatt 1981, 2007, Cartwright 1991, Soler *et al.* 2012

17 Bogen and Woodward 1988, Guala 2005, Aarts *et al*. 2015 pg. 1. See also Bovens and Hartmann 2004 for discussion of reproduction and replication in relation to methods of detection.

18 Guala 2005, pg 13-15, Nosek and Errington 2017.

19 Nosek and Errington 2017 pg. 17.

20 It is not just about what "drives" the results in terms of what experimental conditions, but replication can also reveal if, say, the theoretical interpretation of the results are wrong, but I will put this aside. See Stroebe and Strack 2014 and Nosek and Errington 2017

This uncertainty comes in many forms. There may be concerns about the method of collecting and analyzing data that leads to a false positive. Or, there could be concerns that the experimental conditions played some unexpected role in producing the results.[21] When it comes to uncertainty of results, replication and coherence of evidence generally provide an epistemic boost under many different conditions of uncertainty. For instance, exact replications using the same methods as the initial experiment, also called direct replications, can confirm whether or not the initial experiment produced a false positive (among other things). Divergent replications, where the replication changes some experimental methods, can provide insight into whether or not something about the initial experimental conditions produce the result, focusing on those conditions that are not part of "what is presumed to matter."

The reason there is a concern about experimental conditions influencing the result is because experimental conditions are often quite contrived. The conditions of an experiment are important because they allow the experimenters to try and focus in on what is presumed to matter by removing the buzzing, busy world outside the laboratory and its many confounding factors. There is often a focus on singling out the causal relationship between some limited set of causes in a phenomenon of interest to gain a better understanding of those causal factors, and the experimental conditions may be contrived so as to control the other causal factors found in the target. The purpose of such experiments is to try and determine whether or not there is a causal relationship between some variables, where the conditions that might produce this relationship in the world are too messy. For such an experiment, "what is presumed to matter" is a limited subset of how things function in the world, since the purpose is to discover a causal relationship between a subset of the causally factors. To achieve this goal, experimental conditions are constructed that control, limit, remove, or exaggerate the influence of some causally relevant aspects of the outside world so that a better understanding of the features of interest can be had. In this way they do not reflect the world but can provide insight into some causal relation.

---

[21] This is an inclusive "or". See Cesario 2014, Simons *et al*. 2014, Stroebe and Strack 2014, Maxwell *et al*. 2015.

We can find clear examples of this in nutritional studies. Blood lipid levels in humans are influenced by several factors, only some of which are dietary related and we might be interested in how some particular aspect of diet influences blood lipid levels. For instance, the influence of a particular macronutrient (protein, carbohydrate, or fat) or some subcategory of those (e.g. saturated or unsaturated fat). So, if we are concerned with finding out the influence of saturated fat on HDL cholesterol when compared to unsaturated fat, experiments need to control diets so that subjects are in energy balance (consuming as many calories as they use). Further, the percentages of particular nutrients making up the diet need to be tightly controlled. For instance, in a metabolic ward study, one diet will have 20% of total calories from saturated fat, while a comparison diet may have 10% of calories from saturated fat and replace those lost total calories with a source of unsaturated fat. Such experiments tightly control many causal influences on blood lipid levels that might confound evidence collected from free living populations, such as whether or not a diet is hypercaloric, hypocaloric, or isocaloric, and the activity levels of subjects.

Now there is a fear when introducing such experimental conditions is that the results are not determined by the particular causal relationship of interest, but are influenced in a way that the experimenters do not want by the experimental conditions. Replicating the experiment may involve replacing one way of controlling a certain causal factor for another way of controlling it. In this way it can be determined whether or not that method of control was influencing the results in a way that undercut the scientists' study of what was presumed to matter. The role of replication in experiments includes the ability to determine which causal processes were responsible for the results determined in the initial experiment, and whether or not the results of such an experiment can then lend support to claims about what the scientists took to be what really matters.

In the case of nutrition experiments, foods are not a source of a single nutrient. There is no food that only carries saturated or unsaturated fat, as foods often carry further micronutrients, anti-oxidants, or phytochemicals. Each of these may have an impact on, say, absorption of fat during digestion or production of cholesterol. So, it might be the case that the experimental diets used a source of unsaturated

fat that carried with it some micronutrient that produced some unexpected causal influence, potentially undercutting any inference relating to the impact of saturated or unsaturated fats.

It is important to note that such experimental conditions do not reflect the target as it would be found out in the wild, but are constructed so as to control some of the causal factors to gain a better understanding of those causal influences of interest. More importantly, such experiments can play a confirmatory role, but robustness reasoning is still needed. It is unclear what the underlying causal process is that produced the results of a particular experiment on its own, because the experimental conditions may be playing an unexpected role. However, given the replication of an experiment, or several replications, it can be made clear whether or not the methods of controlling the non-important causally relevant factors played an unwanted influence.

The focus of the above discussion is two-fold. First is that robustness reasoning can still play a distinct confirmatory role even under conditions where an experiment can provide confirmation. Each experiment that is run can provide confirmation of a proposition. However, given a single experiment, it is unclear what causal relation the experiment captures, and it is therefore unclear if the experiment provides confirmation. Given a set of experiments that agree, it can be inferred that each experiment captures "what is presumed to matter" and therefore does provide confirmation.

Further, not representing things as they are outside the laboratory plays an important epistemic role. It allows experimenters to focus in on a causal relation of interest. Given this, an attempt to make the experiment more realistic (i.e. like real-world conditions) would be antithetical to the experimenters goals. The purpose of the contrived experimental conditions is to provide greater clarity on a particular causal relationship. Making the experimental conditions exactly reflect the world would undercut the ability to do so. As I explain below, there is a similar role available for idealizations in experiments, and that de-idealizations would, similarly, be antithetical to the modelers' purposes.

3. Idealizations and Controls

To get my analogy to work, it needs to be shown that models can be understood in terms similar to experimental conditions. To start,that we might want to remove some causal confounders to study some causal relation of interest is not unique to experiments. This is a method common in modeling as well. Elliott-Graves and Weisberg, for instance, discuss what they call "minimal models", where much of the target system is idealized to get a better understanding of the workings of the system.[22]

Michael Strevens has discussed the content of idealizations and distinguishes between the literal content and explanatory content of idealizations.[23] The literal content of an idealization is exactly as it sounds, what the idealization literally implies. The explanatory content of an idealization is more context dependent, and involves the purpose to which the idealization is incorporated into a model. In population biology, for instance, populations might be idealized as infinite.[24] The literal content of this idealization is that a population is infinite and this is obviously false. However, the explanatory content may be that genetic drift is not a causal difference maker. Since genetic drift is related to fluctuations in the frequency of certain alleles in a population due to random chance events, as the population size moves towards the infinite this is able to effectively control this causal influence.

Despite the terminology, we do not need to tie idealizations specifically to explanations *per se*. It might be that scientists are interested in investigating the causal relation between selective pressures and frequency of alleles in the population. To do so, the scientists might want to control the influence of genetic drift by removing its influence altogether, and can do so by representing the population as infinite. So, while there may be no actual populations that completely lack the causal influence of genetic drift, controlling the influence of this in our investigations can provide greater insight into other causal relations we are interested in. Just like experimental controls remove causal confounders, idealizations can also play this role.

---

22 Elliott-Graves and Weisberg 2014

23 Strevens 2008, ch 8., 2013.

24 As has been pointed out, Gillespie 1998, Strevens 2018, populations are not actually infinite in these models, but are more asymptotic or moving towards a limit.

Now, given this account of idealizations as experimental controls, we can extend the analogy of robustness reasoning in experiments to that of models. Given the construction of a model using one idealization as a means of control, there is a concern that this idealization might be influencing the results of the causal relations we are concerned with in a way that is unknown. So, while the model is able to confirm some claim about causal relations, it is unclear if the model is actually capturing the causal relation of interest. So, a model with different sets of idealizations may be constructed to show that the initial means of control is not influencing the results, and allow the scientists to confirm whether or not the causal relation of interest is driving the result.

If the analogy between experiments and models holds, then this carves out a confirmatory space for robustness analysis when applied to models. Idealizations can play the role of experimental controls and robustness analysis is a means of showing that these controls are effective, providing insight into the causal relation of interest.

4. Concessions and Objections

There are several concerns to be answered about this account. First, I will make some concessions. I think critics of robustness analysis are correct that many cases where robustness analysis has been applied do not carry the confirmatory power they are thought to.[25] Even if we accept that idealizations *can* play the role of controlling conditions, this does not mean that this is always the role that they play. There are cases where robustness analysis has been attempted where the idealizations might not play this role of control. However, that this criticism holds in many actual cases does not mean that it holds generally.

This concession is fairly significant, however, since viewing idealizations as means of control and the particular analogy between robustness and replication that I draw is not held. To be sure, the comparison between robustness in experiments and robustness in models has been discussed, but this

[25] Odenbaugh and Alexandrova 2011 provide some clear examples

does not mean that robustness is often treated this way.[26] In the example drawn from Levins, three models were constructed to test the robustness of a theory. All of these models were idealized, but some traded out one idealization for a de-idealization. The optimality models, for instance, did not incorporate any genetic information and we might consider this a way of controlling genetic influence on the phenotypes of interest so that selection pressures could be better studied. Optimality models were traded out for a genetic model to complete Levins' robust set and he takes their agreement to confirm the shared robust theorem.

This is slightly different from how I view robustness. Although the genetic model may be capturing the same "core biological assumption" according to Levins, it is capturing a different core causal mechanism. The two optimality models, in so far as they trade out idealizations that might be interpreted as means of controlling causal influences, are more in-line with the view of robustness I have in mind. This has further effects on the understanding of robustness analysis as the robust theorems that can be confirmed will need to be altered. The more general claim that Levins makes about species in uncertain environments would need to be indexed to a more specific claim about, say, selection pressures tending towards broad niches in uncertain environments.[27]

To return to our example from nutritional studies examining the influence of particular dietary fats (saturated and unsaturated) on blood lipid levels, it would not be helpful in testing this to run an experiment that included an exercise component as this introduces a new causal component. Similarly, a robust set of models should focus on a particular causal relation.

These concessions are fairly significant since it paints a somewhat different picture of robustness analysis than is common. However, it is one that may have some plausibility, particularly given some discussion below, and can open up room for robustness analysis to play a role in confirmation.

---

[26] See Cartwright 1991 and Odenbaugh and Alexandrova 2011 for ways these analogies might fail.

[27] It is still possible to bring together genetic and optimality models to make this more general claim. However, as I view it, there would need to be separate robustness analysis on the specific causal mechanisms captured in the model. So, a robustness analysis on selective pressures and a robustness analysis on genetic influences. If both of these agree, as Levins takes to be the case, I take this to be two bits of information in favor of Levins' more general claim and would count more in terms of accumulation of evidence.

There are several important concerns relating to the ontological differences between models and experiments.[28] Models being mathematical and experiments being physical does provide some important epistemic distinctions. For instance, the causal relation being studied is not part of the mathematical framework the way that it is part of an experimental subject and so this question about a model actually specifying a causal relation is better motivated.[29]

While this is an important point, it also needs to be noted that the fact that the causal relation is built into the experimental subject does not mean that models cannot capture all that is needed to appropriately represent the causal relation of interest. It does point out an important difference in degree, in that we have more reason to believe that our experimental subjects instantiate the causal relation of interest, but this difference in degree is not insurmountable. We might have good reason to believe that our mathematical framework is up to the task of representing the causal relation of interest.[30]

As noted, this is a difference in degree, as there is an analogous assumption when it comes to experimentation. We need to have some reason to believe that the methods of measurement being employed are up to the task of detecting the causal relation in question, even if this question does not arise in the experimental subjects. Perrin famously used multiple experiments to calculate Avogodro's number to determine the existence of atoms. This replication would only be useful if calculating Avogodro's number was a worthwhile measurement for determining the existence of atoms. So, while it might be that the causal relation is necessarily a part of the physical experiment in a way that it is not part of a mathematical model, this does not go so far as to present an insurmountable problem. The clarification needed is simply that we need some reason to believe that our methods of representing and detecting the causal relation are up to the task, in both models and experiments.

There are two further concerns derived from this ontological difference (also pointed out by Morgan, 2012, chap. 6.6). One difference between models and experiments is that, despite all of our best

---

[28] See Maki 2005, Morgan 2012, Ch. 6 section 6, and Morrison 2015 for more discussion of the epistemic distinctions and similarities between models and experiments.

[29] See Humphreys 2002, Alexandrova 2008, and Odenbaugh and Alexandrova 2011 for more on this concern.

[30] See Weisberg 2006.

efforts, we are limited in how much we can engineer experimental conditions. It is beyond our abilities, for instance, to introduce impossible experimental conditions in the way that we can in models. This difference, however, is not a problem as long as we keep in mind that the idealizations we are considering are those that control some causally relevant factors that we are interested in controlling. Whether this is done by impossible or possible means is a matter of degree in this way, and both can be shown to *not* affect the results detected in a similar manner.

Further, this ontological difference carries with it epistemic advantages of sorts for models. Morgan distinguishes between "surprise" and "confoundment". Models are able to surprise scientists with the results produced but not confound, while physical experiments can confound. Since models are constructed, all of the elements that produce the result are known, but it is unknown how the combination of variables work together. So, scientists might be surprised by a result of a model, but it is never due to some unknown variable in the model, since all of the variables were put there. Physical experiments, on the other hand, have the potential to confound. Since knowledge of all of the factors that might play out in a physical system is limited, it is possible for the results of an experiment to present something legitimately new that is not explained or part of current theory. Models, since all of the parts are known, are not able to produce this novel experience. This is to the advantage of models in terms of focusing in on relevant causal factors. There is no possibility for confoundment, since all of the parts of the model are known because they were explicitly put in there, if our interest is to zero in on some causal relation, models can prove to be a better option in some cases.

Another line of criticism might stem from the distinction between literal and explanatory content and whether or not it is *ad hoc*. As with many things in the philosophy of science, an account of idealizations needs to answer not only philosophical concerns but also to the practices of science. Part of why idealizations have been so philosophically vexing is that they seem to be epistemically corrupting but are both prevalent and stubborn in scientific practice. Idealizations are commonplace but it seems that

de-idealization is not often a goal.[31] An account that can remove part of this tension has some *prima facie* support.

This distinction between literal and explanatory content can provide some answers to these concerns. While a complete defense is beyond the scope of this paper, it can help make sense of the prevalence and stubbornness of idealizations. Importantly, providing controls on causal influences is a common part of scientific investigations, and if this can be extended to an account of modeling (at least in some instances) this is an advantage. Further, that idealizations might be introduced as a control does not conflict with other reasons we might use to explain the prevalence of idealizations in models, such as pragmatic reasons. We can introduce idealizations for a variety of epistemic reasons, and it might be that some idealizations are introduced for pragmatic or simplicity reasons sometimes, while other times they are introduced for reasons of control.

Further, this distinction provides one way of making sense of why de-idealization is rarely the end-goal of a model.[32] While models are often traded out or altered, it is not always the case that a model is headed toward a de-idealized state. Rather than iteratively "improving" models by de-idealizing as we learn more, idealized models are sometimes kept more or less the same. Once again, it might be that idealizations are kept for pragmatic reasons, where the model would just be too complex if de-idealized, but this is not always the case. That idealizations play a positive epistemic role, not just pragmatic, can provide some insight into why de-idealization is not always the goal. This lends some, at least *prima facie*, reason to accept the distinction between literal and explanatory content.[33]

Given the considerations above, we can acknowledge that there are distinctions between models and experiments. However, when it comes to the robustness I have been discussing, these differences

---

31 Levins 1966, Morrison 2015, Rice 2019, 2021.

32 As Levins (1966, pg. 430) points out, it is rare that a model is made completely precise, but often they are simply replaced with a different model that makes its own idealizations.

33 It could be argued that idealizations are kept for pragmatic and reasons of simplicity. However, if idealizations were seen as providing a legitimate obstacle to understanding it seems that they would be de-idealized. Further, idealizations could be dealt with as computational power increases. It is the case that models have gotten more complex with the incorporation of more computational power into scientific research. However, this has not resulted in de-idealized models, but more complex models. For instance, ecology has started incorporating Individual Based Models, which are not de-idealized versions of older models that focused on populations, but a different and more complex type of model.

often amount to differences in degree and not a difference in degree extreme enough to warrant legitimate concern about the possibility of confirmation by robustness.

5. Conclusion

The above discussion has an upshot, but also makes some considerable concessions. Most significantly, I agree with critics of robustness analysis that many actual cases of robustness analysis on models do not carry the confirmation attributed. At best, it is often unclear what confirmatory support is provided by robustness analysis. A further concession is that the conditions under which robustness analysis might be applied are somewhat constrained. If the idealizations are not introduced with an understood explanatory content, or if they are not introduced with the intention of controlling some causal influence, then what I have said lends little support.

The upshot is that there may be a role for robustness in confirming models. Further is that introducing idealizations as a means of control by employing some understood notion of explanatory content is not an improbable use for idealizations. Further still, it seems that this means of focusing in on certain causal influences is not a rare or unheard of use for models, and so this is a plausible and worthwhile use of robustness analysis.

Works Cited

1. Aarts, A.A, et al. “Estimating the Reproducibility of Psychological Science.” Science (American Association for the Advancement of Science), vol. 349, no. 6251, 2015, p. 943.
2. Achinstein, Peter. The Book of Evidence. Oxford University Press, 2001.
3. Alexandrova, Anna. “Making Models Count.” Philosophy of Science, vol. 75, no. 3, 2008, pp. 383–404.
4. Bogen, J, and J Woodward. “Saving the Phenomena.” The Philosophical Review, vol. 97, no. 3, 1988, pp. 303–352.

5. Cartwright, Nancy. “Replicability, Reproducibility, and Robustness: Comments on Harry Collins”. History of Political Economy, vol. 23, no. 1, 1991, pp 143-155.
6. Cesario, Joseph. "Priming, Replication, and the Hardest Science : Behaviorial Priming and Its Replication." *Perspectives on Psychological Science* 9, no. 1, 2014, 40-48.
7. Elliott-Graves, Alkistis, and Michael Weisberg. “Idealization.” Philosophy Compass, vol. 9, no. 3, 2014, pp. 176–185.
8. Guala, Francesco. The Methodology of Experimental Economics. Cambridge University Press, 2005.
9. Humphreys, Paul. “Computational Models.” Philosophy of Science, vol. 69, no. S3, 2002, pp. S1–S11.
10. Kuorikoski, Jaakko, et al. “Economic Modelling as Robustness Analysis.” The British Journal for the Philosophy of Science, vol. 61, no. 3, 2010, pp. 541–567.
11. Nosek, Brian A, and Timothy M Errington. "Making Sense of Replications." ELife, Vol 6., 2017.
12. Richard Levins. “The Strategy of Model Building in Population Biology.” American Scientist, vol. 54, no. 4, 1966, pp. 421–431.
13. Levins, Richard. “A Response to Orzack and Sober: Formal Analysis and the Fluidity of Science.” The Quarterly Review of Biology, vol. 68, no. 4, 1993, pp. 547–555.
14. Maaki, Uskali. “Models Are Experiments, Experiments Are Models.” The Journal of Economic Methodology, vol. 12, no. 2, 2005, pp. 303–315.
15. Maxwell, Scott E, Michael Y Lau, and George S Howard. "Is Psychology Suffering From a Replication Crisis?" *The American Psychologist* 70, no. 6, 2015, pp. 487-98.
16. Morrison, Margaret. Reconstructing Reality : Models, Mathematics, and Simulations. 2015.
17. Morgan, Mary S. The World in the Model. Cambridge University Press, 2012.
18. Odenbaugh, Jay, and Anna Alexandrova. “Buyer Beware: Robustness Analyses in Economics and Biology.” Biology & Philosophy, vol. 26, no. 5, 2011, pp. 757–771.

19. Orzack, Steven Hecht. “Discussion: What, If Anything, Is ‘The Strategy of Model Building in Population Biology?’ A Comment on Levins (1966) and Odenbaugh (2003).” Philosophy of Science, vol. 72, no. 3, 2005, pp. 479–485.
20. Orzack, Steven Hecht, and Elliott Sober. “A Critical Assessment of Levins's The Strategy of Model Building in Population Biology (1966).” The Quarterly Review of Biology, vol. 68, no. 4, 1993, pp. 533–546.
21. Rice, Collin. "Models Don't Decompose that Way: A Holistic View of Idealized Models", The British Journal for the Philosophy of Science, vol. 70, no.1, 2019, pp. 179-208.
22. Rice, Collin. Leveraging Distortions : Explanation, Idealization, and Universal Patterns in Science. 2021.
23. Schulz, Armin W. "The Heuristic Defense of Scientific Models: An Incentive-Based Assessment." *Perspectives on Science* 23, no. 4, 2015, pp. 424-42.
24. Simons, Daniel J. "The Value of Direct Replication." *Perspectives on Psychological Science* 9, no. 1, 2014, 76-80.
25. Stroebe, Wolfgang, and Fritz Strack. "The Alleged Crisis and the Illusion of Exact Replication." *Perspectives on Psychological Science* 9, no. 1, 2014, 59-71.
26. Sugden, Robert. “Credible Worlds: the Status of Theoretical Models in Economics.” The Journal of Economic Methodology, vol. 7, no. 1, 2001, pp. 1–31.
27. Weisberg, Michael. "Robustness Analysis." *Philosophy of Science* 73, no. 5, 2006, pp. 730-42.
28. Wimsatt, William C. “Robustness, Reliability, and Overdetermination.” in Soler *et al.* (eds). Characterizing Robustness in Science: After the Practice Turn in the Philosophy of Science. Springer-Verlag, 2012. Pp. 61-78.
29. Wimsatt, William C. Re-Engineering Philosophy for Limited Beings : Piecewise Approximations to Reality. Harvard University Press, 2007.

30. Woodward, Jim. “Some Varieties of Robustness.” The Journal of Economic Methodology, vol. 13, no. 2, 2006, pp. 219–240.