Robustness and Independent Evidence

Jacob Stegenga and Tarun Menon*†

Robustness arguments hold that hypotheses are more likely to be true when they are confirmed by diverse kinds of evidence. Robustness arguments require the confirming evidence to be independent. We identify two kinds of independence appealed to in robustness arguments: ontic independence (OI)—when the multiple lines of evidence depend on different materials, assumptions, or theories—and probabilistic independence. Many assume that OI is sufficient for a robustness argument to be warranted. However, we argue that, as typically construed, OI is not a sufficient independence condition for warranting robustness arguments. We show that OI evidence can collectively confirm a hypothesis to a lower degree than individual lines of evidence, contrary to the standard assumption undergirding usual robustness arguments. We employ Bayesian networks to represent the ideal empirical scenario for a robustness argument and a variety of ways in which empirical scenarios can fall short of this ideal.

1. That Special Epistemic Oomph. Many suppose that when a variety of evidence supports a hypothesis, this hypothesis is more likely to be true. The property of inductive arguments that licenses this supposition is sometimes called 'robustness' (we will call such arguments 'robustness arguments'). Philosophers have employed robustness arguments as support for various forms of realism, including causal realism, entity realism, natural kind realism, and theory realism and to counter the experimenter's regress argument, to resolve skeptical underdetermination concerns, to distinguish artifacts

Received December 2015; revised September 2016.

*To contact the authors, please write to: Jacob Stegenga, Department of History and Philosophy of Science, University of Cambridge, Cambridge CB2 3RH, United Kingdom; e-mail: jms303@cam.ac.uk. Tarun Menon, Tata Institute of Social Sciences, Mumbai, India; e-mail: tarunium@gmail.com.

†For commentary and discussion we thank Chiara Lisciandra, Aki Lehtinen, Jonah Schupbach, Kent Staley, Jaakko Kuorikoski, Caterina Marchionni, Eric Barnes, and audiences at the American Philosophical Association, the Philosophy of Science Association, and the 2014 workshop on robustness in Helsinki. Support for Stegenga was provided by the Social Sciences and Humanities Research Council of Canada.

Philosophy of Science, 84 (July 2017) pp. 414–435. 0031-8248/2017/8403-0002$10.00
Copyright 2017 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/692141 Published online by Cambridge University Press

from real entities, to aid us in our pursuit of objectivity, and to permit the observation of unobservable entities.

As an argument for one of these forms of scientific realism, robustness is often formulated as a no-miracles argument, or an argument from coincidence. Here, for example, is Salmon (1997) discussing the canonical Perrin episode, based on the agreement of multiple methods of measuring Avogadro's number: "such agreement would be miraculous if matter were not composed of molecules and atoms." That is, it would be a miracle if diverse evidence supported a hypothesis and the hypothesis were not true; we do not accept miracles as compelling explanations; thus, when diverse evidence supports a hypothesis, we have strong grounds to believe that it is true. It is often assumed that diverse evidence that supports a hypothesis provides a special epistemic oomph to the hypothesis.[1] In what follows we formulate the notion of diversity of evidence and show that only very particular empirical scenarios warrant a robustness argument, and such scenarios are constrained in ways not usually recognized in the wide literature that appeals to robustness arguments.[2]

Robustness arguments are said to be compelling only if the multiple kinds of available evidence are independent. Regarding Perrin's multiple methods of measuring Avogadro's number, Salmon (1984) urges his reader to "notice what a wide variety of substances are involved and how diverse are the phenomena being observed." Culp (1995) considers robustness arguments compelling if the following condition is met: "the techniques must not all use the same theoretical presuppositions in making raw data interpretations" (450). Similarly, when Hacking (1983) asked "Do we see through a microscope?" he was interested in the experimental strategies that a scientist could use to be confident that features of an object observed under a microscope are real

1. Among many others, see Cartwright (1983), Salmon (1984), Bechtel (2002), Snyder (2005), and Kuorikoski and Marchionni (2016). Robustness was famously discussed in the nineteenth century by Whewell (1837), who gave it the name "consilience of inductions." Sometimes this inductive principle is held to be axiomatic: "It is an analytic proposition . . . that, other things being equal, the evidence for a generalization is strong in proportion as the number of favorable instances, and the variety of circumstances in which they have been found, is great" (Strawson 1952, 256–57).

2. Our concern is with concordant evidence for an empirical hypothesis. There is a growing literature on robustness of models. For example, Levins (1966, 420) and Wimsatt (1981) have argued that robustness is valuable for modeling—because models employ idealizations, "truth is the intersection of independent lies." In contrast, Cartwright (1991) argued that robustness arguments in econometric modeling are crude inductions, and Orzack and Sober (1993) have criticized model robustness on the charge of its being a nonempirical form of confirmation. The debate continues today by, e.g., Weisberg (2006), Kuorikoski, Lehtinen, and Marchionni (2010), Odenbough and Alexandrova (2011), and Parker (2011). The present article is not concerned with robustness of models.

https://doi.org/10.1086/692141 Published online by Cambridge University Press

and not artifacts of the apparatus. Confidence is gained, according to Hacking, if different kinds of microscopes are used, because "these processes have virtually nothing in common between them. They are essentially unrelated chunks of physics." The independence condition is meant to ensure that the concordant evidence from multiple methods is due to the object of investigation rather than an error-prone feature shared by the methods. It is independence of the material or theoretical basis of observational apparatuses—what we call *ontic* independence (OI)—that many philosophers have assumed to be the relevant kind of independence required for robustness arguments. Substances, theories, or chunks of physics: these are the sorts of things that are often said must differ between lines of evidence in order to satisfy the independence condition for robustness arguments. When ontically independent evidence supports a hypothesis, it is often thought that one can construct a robustness argument.[3]

In order to explore the consequences of evidential independence and to represent the various ways in which such independence can fail, we employ the graphical device of Bayesian networks.[4] We use Bayesian networks first to represent an empirical scenario that exemplifies an ideal robustness argument (sec. 2). The ideal robustness argument satisfies ontic independence but, more importantly, satisfies what we call conditional probabilistic independence (CPI). The Bayesian network representation allows us to show that in such ideal empirical scenarios, adding different kinds of independent evidence is guaranteed to increase confirmation, thereby providing some warrant to robustness arguments. We then use a Bayesian network to represent the usual way in which robustness is said to fail, namely, when the various kinds of evidence are not ontically independent (sec. 3). Two short case studies illustrate such a scenario. Even when the various kinds of evidence are ontically independent and all the evidence confirms the same hypothesis, a robustness argument may not be warranted. A radical way robustness arguments can fail to be warranted even when the evidence is ontically independent involves scenarios in which the evidence is 'dyssynergystic': again using Bayesian networks, we represent possible empirical scenarios in which this can happen (sec. 4). If the different kinds of evidence are CPI, however, then we can show that the evidence is not dyssynergystic. We conclude by discussing logical and epistemic constraints on robustness arguments (sec. 5).

3. Hudson (2014) offers a rare criticism of the assumption that converging OI evidence warrants robustness arguments. Though our conclusion is broadly consistent with Hudson's, our approach is quite different; specifically, we are unsatisfied with his treatment of probabilistic independence, but arguing the point here would take us astray.

4. Bovens and Hartmann (2004) have led the way for this approach, with their insightful analysis of evidence and confirmation using Bayesian networks. Claveau (2013) has more recently developed the approach of Bovens and Hartmann.

https://doi.org/10.1086/692141 Published online by Cambridge University Press

In short, robustness arguments require different kinds of evidence to be independent, and explications of this notion of independent evidence fall into two families: OI and CPI. Many strong conclusions have been drawn on the basis of robustness arguments in which the evidence is OI, and we show that such arguments are not generally justified. As typically construed, not all OI evidence is CPI. However, OI and CPI are not necessarily distinct: employing the graphical device of Bayesian networks, we show one way in which OI evidence can be structured in an empirical scenario such that the conditions of CPI are satisfied and thus that one is ensured an increase in confirmation of one's target hypothesis.

2. Ideal Robustness Arguments. We take a very general view of what must be robust in robustness arguments, namely, confirmational support that diverse lines of evidence provide to a hypothesis. In the Perrin case so often discussed by philosophers, it was the measurement of a physical quantity (Avogadro's number) that was robust. But robustness can be characterized in a more general way than agreement among measurements of a quantity by different lines of evidence: robustness can involve multiple lines of evidence providing confirmational support to the same hypothesis. Our cases described in section 3 are exemplary in this regard. Woodward (2006) offers a catalogue of various kinds of robustness—including measurement robustness, derivational robustness, causal robustness, and inferential robustness—and suggests that considerations that undergird one type of robustness may not undergird another type of robustness. We think it worthwhile to attempt to formulate an ideal robustness argument, the logic of which is general with respect to undergirding different kinds of robustness. That is what follows here.

To represent the ideal robustness argument we employ Bayesian networks. A Bayesian network is a probabilistic graphical model. We start with a set of variables and a joint probability distribution defined over them. The structure of conditional dependencies between these variables is represented in a directed acyclic graph, where the variables are the nodes and the connections or edges represent the dependencies.[5] If there are two nodes A and B such that $\Pr(A|B) \neq \Pr(A)$, and this correlation is not screened off by conditionalizing on all of A's parents in the graph (excluding B), then there is a connection between A and B.[6]

5. A graph consists of nodes with connections between them. A graph is directed if the connections have directions associated with them. A graph is acyclic if there are no loops from a variable back to itself. We use capital letters to represent variables and italics to represent values of those variables; thus X represents the variable X, which in the binary case can take the value X (X is true) or $\sim X$ (X is false).

6. For more on Bayesian networks and their application in causal discovery, see Pearl (2000).

https://doi.org/10.1086/692141 Published online by Cambridge University Press

The dependency relations represented by our Bayesian networks are inferential. Of course, inferential dependencies might arise as a result of causal dependencies. To use an example of Bovens and Hartmann (2004), suppose a disease is characterized as the presence of bacteria in the blood (R1), and there are two symptoms that accompany this disease: internal bleeding (R2) and inflammation of the lymph system (R3). Characteristic R1 has a "direct influence" on R2 and R3, to use the terminology of Bovens and Hartmann (2004, 71). Of course, from R2 and R3 we can infer R1 (a diagnostic inference). Or from R1 we could infer R2 and R3 (a prognostic inference). Neither of the symptoms has an influence on each other: R2 is independent of R3 once we conditionalize on R1. In any case, it is the probabilistic relations that permit the inferences, and it is the causal influence from R1 to R2 and R3 that determines the directionality of the edge in their graph that represents this case.

For our purposes we need only one rule about inferring dependencies from graphical structure. Articulating the rule, however, will require some terminology. Let X and Y be two variables in a graph **G**, and **Z** be a set of variables in **G** that does not contain either X or Y. A *trail* between two variables is an undirected path between those two variables, that is, a path that follows connections, but not necessarily in the direction of those connections. A variable C on a trail is a *collider* relative to that trail if and only if it is not one of the end points of the trail and both the incoming and outgoing connections to C on that trail are directed toward C. So the trail near C can be represented by figure 1.

D-separation is a central notion for Bayesian networks and can be used to determine if two variables X and Y are independent, given **Z**. Variables X and Y are d-separated by **Z** if and only if there is at least one variable on every trail between X and Y that is either not a collider and is in **Z** or is a collider and is neither in **Z** nor is an ancestor of any variable in **Z**. This condition ensures that the *Markov condition* holds. The Markov condition entails that variables X and Y are probabilistically independent conditional on **Z** if and only if **Z** d-separates X and Y. This allows one to visually determine conditional (in)dependencies in a Bayesian network.

In figure 2, H is the hypothesis of interest, and C1 and C2 are two consequences of this hypothesis that are open to empirical test. These need not be deductive consequences: it may be the case that $\Pr(C1|H) < 1$. However,

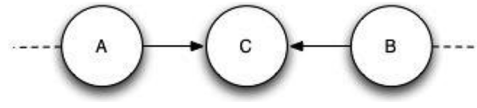


Figure 1. C is a collider on the trail between A and B.

https://doi.org/10.1086/692141 Published online by Cambridge University Press

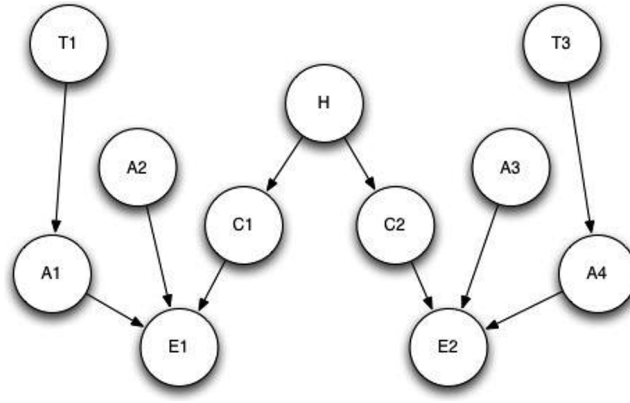


Figure 2. Bayesian network representation of an ideal robustness argument.

we do require that $\Pr(C1|H) > \Pr(C1)$. The variables E1 and E2 represent evidential states that indicate whether or not C1 and C2, respectively, are true or false and thereby are evidence relevant to the truth or falsity of H. In other words, we have

$$\Pr(E1|C1) > \Pr(E1).$$

The pieces of evidence are indicators of the truth or falsity of H only if certain auxiliary assumptions about the observational apparatus and empirical scenario hold (represented by variables A1–A4). The auxiliary assumptions are supported by background theories (represented by variables T1 and T3).

The two pieces of evidence E1 and E2 may be influenced by auxiliary assumptions (A2 and A3) that rely on the same background theory, which would apparently violate the premise of robustness arguments that the modes of evidence must be independent. We doubt that we are ever in an empirical situation in which the modes of evidence rely on entirely distinct sets of background theories. In all real cases there will be at least some common theoretical background to the multiple modes of evidence. Despite that, if we are sufficiently convinced of such theories, they ought not be considered a threat to robustness arguments. For instance, if the common assumptions stem from a highly confirmed physical theory, such as wave optics, they do not engender a problematic form of dependence between the modes of evidence. Problems for robustness arise if the common background assumptions or associated theories are themselves controversial. Intuitively, if there is a nonnegligible chance that we are wrong about a background assumption, then we must consider the possibility that concordance among modes of evidence that share this assumption is attributable to the false assumption rather than the truth of the investigated hypothesis. However, this

https://doi.org/10.1086/692141 Published online by Cambridge University Press

is not a credible problem if the common assumption is extremely well substantiated. In our ideal case, there may be a common theory legitimating the particular auxiliary assumptions. However, if this theory is sufficiently well established that we can treat it as essentially having probability one, then it drops out of the graph as a relevant factor.

In figure 2 all the auxiliary assumptions are indeed unconditionally independent. The theories T1 and T3 are not common assumptions of both modes of evidence. These assumptions might be controversial. As above, robustness does not require the independence of all theoretical assumptions between the modes of evidence; rather, it requires only independence of problematic or controversial auxiliary assumptions.

The ideal robustness argument can be validated by appeal to figure 2. The hypothesis H d-separates E1 and E2.[7] The only trail between these variables is E1-C1-H-C2-E2, and H is on this trail and not a collider. It follows that

$$\Pr(E1 \,\&\, E2|H) = \Pr(E1|H) \times \Pr(E2|H).$$

We have already stipulated that $E1$ and $E2$ individually confirm H. This means that the likelihood ratios associated with these pieces of evidence must be greater than one. The likelihood ratio for $E1$ is

$$\mathrm{LR}_{E1} = \Pr(E1|H)/\Pr(E1|{\sim}H) > 1.$$

A similar expression holds for LR_{E2}, the likelihood ratio for $E2$. The independence of $E1$ and $E2$ conditional on H allows us to calculate the likelihood ratio for the conjunction of both pieces of evidence:

$$\begin{aligned}\mathrm{LR}_{E1\&E2} &= \Pr(E1 \,\&\, E2|H)/\Pr(E1 \,\&\, E2|{\sim}H)\\ &= \mathrm{LR}_{E1} \times \mathrm{LR}_{E2}.\end{aligned}$$

The second equality follows from the independence assumption. Since LR_{E1} and LR_{E2} are both greater than one, it follows that $\mathrm{LR}_{E1\&E2}$ must be greater than either of the individual likelihood ratios (this argument appears in Sober [2008]). A higher likelihood ratio corresponds to a higher degree of confirmation, so the conjunction of the evidence is more confirmatory than either individual piece of evidence.[8] Here, then, we have some grounds for the de-

7. More precisely, the set of variables {H} d-separates E1 and E2. However, in what follows we will leave out the braces when speaking of a singleton set, since nothing in our argument hinges on distinguishing a singleton set and its element.

8. The posterior odds of a hypothesis are obtained by multiplying the prior odds by the likelihood ratio; so given identical prior odds, a greater likelihood ratio will raise the probability of the hypothesis to a greater degree. This argument demonstrating the increase in confirmation based on diverse evidence is not sensitive to the use of the likelihood ratio as a measure of confirmation; see the appendix for a general proof.

https://doi.org/10.1086/692141 Published online by Cambridge University Press

sirability of independent evidence. If we can ensure that the evidence we collect will be independent conditional on the truth of the hypothesis being investigated, then we know that the more confirmatory evidence we gather, the more the hypothesis is confirmed.

3. Failure of Ontic Independence. Most philosophical discussions of robustness have presumed that one needs only to ensure that the modes of evidence themselves are independent. The focus has been on departures from the ideal case represented by figure 3. In figure 3 there is a trail E1-A2-E2 that does not contain H. Hypothesis H no longer d-separates E1 and E2, and so the warrant for the ideal robustness argument no longer applies. This is a way of representing the fact that E1 and E2 are not OI.

If various kinds of evidence support a hypothesis, and thus seem to warrant a robustness argument, yet the kinds of evidence are not OI, then the inductive argument has the property of 'pseudorobustness'.[9] The following examples are illustrative.

3.1. Mesosomes. In the early days of electron microscopy, multiple methods of preparing microscope samples suggested the existence of an as-yet undiscovered cellular organelle, dubbed the 'mesosome', an entity that is now considered an artifact (Rasmussen 1993). During the 1950s and 1960s, cell biologists claimed that they had discovered this new structure within *Bacillus* cells. Fitz-James (1960) called such structures mesosomes and suggested that they were organelles of many bacteria. Fitz-James used multiple fixation techniques to preserve the bacteria and observed mesosomes using both electron microscopes and light microscopes. Fitz-James also claimed to observe mesosomes in living bacteria by using a special staining technique. Thus it appears that a robustness argument could have been formulated for the existence of mesosomes. This is precisely what Fitz-James did.[10] The warrant for such a robustness argument continued to get stronger. A large body of work, which continued into the 1970s, was concerned with determining the function of the mesosome. Biochemists purified mesosomes and subjected them to numerous tests, and this was done with multiple species of bacteria.

9. The term 'pseudorobustness' was introduced, as far as we can tell, by Wimsatt (1981).

10. Fitz-James made a robustness argument despite some apparent discordance between the kinds of evidence. Rasmussen (1993) argued that concordant evidence was used as an argument for robustness, but discordance was explained away by appealing to differences in experimental setup. See also Culp (1995).

https://doi.org/10.1086/692141 Published online by Cambridge University Press

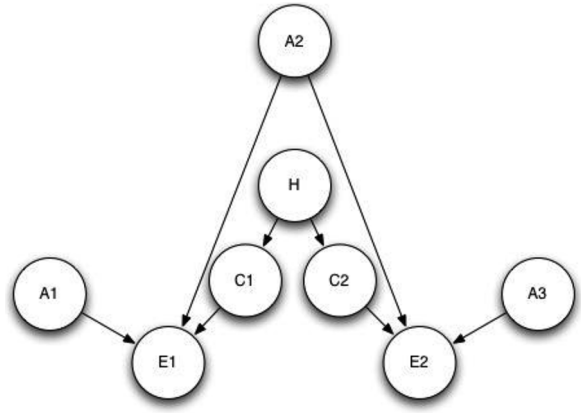


Figure 3. Auxiliary assumption (A2) influences both modes of evidence, and so H no longer d-separates E1 and E2.

Ultimately most scientists came to think that mesosomes are artifacts, and this shift was partially due to new methods of fixing bacteria.[11] Thus from what we know now, this is a case of pseudorobustness. The modes of evidence presented by Fitz-James relied on a shared technique for preparing the microscope samples. With respect to a robustness argument, this means that there was an inferential bottleneck like that of A2 in figure 3. The available evidence was not in fact sufficiently independent for a robustness argument to be justified; that is, the available evidence was not OI.

If one wants to justify a robustness argument, then an obvious methodological prescription is to avoid such inferential bottlenecks. The difficulty with this methodological prescription is that one must know that the various kinds of evidence are actually OI: the presence of inferential bottlenecks must be discernible. Sometimes this is simple. The following case is another example of pseudorobustness in which it was indeed possible to identify inferential bottlenecks. Moreover, although the mesosome case was an example of a pseudorobustness argument supporting a false hypothesis, not all cases of pseudorobustness are such that the apparently diverse evidence supports a false hypothesis. The following is a case in which an inductive argument for a hypothesis was criticized as pseudorobust but later that hypothesis was accepted as true.

11. Hudson (1999) argued that this suggests that the relevant scientists were less swayed by the concordance of diverse evidence than they were by evidence from a single mode that was deemed more reliable than other modes. Hudson calls this 'reliable process reasoning'. See also Hudson (2014).

https://doi.org/10.1086/692141 Published online by Cambridge University Press

3.2. The Chemical Composition of Genes. Determining the material basis of heredity was an important research program in the first half of the twentieth century. The evidence presented in the famous paper by Avery, Macleod, and McCarty (1944) is a paradigm example of pseudorobustness, about a hypothesis now widely thought to be basically correct.

A phenomenon called 'transformation' of bacteria motivated this work. One type of bacteria (nonvirulent, morphologically rough pneumococci) could be transformed into another type (virulent and morphologically smooth): heat-killed smooth virulent pneumococci could be injected into mice along with live nonvirulent rough pneumococci, and the live bacteria would change virulence and morphology (from nonvirulent to virulent, and from rough to smooth). This phenomenon of transformation was possibly a kind of hereditary phenomenon. Thus it was thought that the chemical basis of the transformation substance (TS) could have genetic significance.

It was assumed by most that the TS was a protein: proteins were known to be highly variable, whereas nucleic acids were thought to be a repetitive structural molecule. The structure of TS was assumed to be complex, because the phenotypic features transferred between pneumococcal types were complex: a complex function, it was thought, must be caused by a complex structure.

Avery et al. (1944) provided evidence that the TS is composed of DNA. A reconstruction of the hypothesis and evidence of Avery et al. is as follows:

Hypothesis H. The molecule that causes transformation (the TS) is DNA.

Evidence:

Mode 1. Chemical analysis of TS: (E1): the amounts of carbon, hydrogen, nitrogen, and phosphorous were close to the theoretical values for DNA.

Mode 2. Effect of protein and ribonucleic acid–degrading enzymes on TS: (E2): protein and ribonucleic acid–degrading enzymes had no effect on TS.

Mode 3. Effect of DNA-degrading enzyme on activity of TS: (E3): DNA-degrading enzyme inactivated the TS.

Mode 4. Ultraviolet absorption of TS: (E4): ultraviolet absorption of TS was characteristic of DNA.

Mode 5. Electrophoretic movement of TS: (E5): electrophoretic movement of TS was characteristic of DNA.

Mode 6. Molecular weight analysis of TS: (E6): molecular weight of TS was characteristic of DNA.

https://doi.org/10.1086/692141 Published online by Cambridge University Press

Thus, seemingly diverse evidence supported H, apparently warranting a robustness argument.

Critics pounced. The main experimental concern was that the TS was very likely impure and could have had trace amounts of protein in it. It was this protein, critics claimed, that was the true cause of transformation. The chemical tests available at the time were not sensitive enough to detect the presence of up to 5% protein, and the enzymatic experiments could have been ineffective in degrading an active protein, especially if it was covered by structural nucleic acids (for more details on this case, see Stegenga [2011]).

Not only did this criticism imply that any particular piece of evidence might have been an artifact, the criticism also implied that every kind of evidence used by Avery et al. (1944) shared an inferential bottleneck: the isolation and purification of the TS. All of modes 1–6 were a form of analysis on the TS and as such relied on the same method of isolating and purifying TS; thus, the evidence was not OI. In this case, this inferential bottleneck was thought by critics to be unreliable.

3.3. Summary. Inductive arguments with the property of pseudorobustness are sometimes mustered to support what later comes to be seen as a false hypothesis, as in the mesosome case, but are sometimes mustered to support what later comes to be accepted as a true hypothesis, as in the DNA case. Since robustness is widely thought to be a virtue of inductive arguments, then, when diverse evidence is available for a hypothesis, to avoid the kind of pseudorobustness discussed in this section one must ensure that the various kinds of evidence appealed to are OI.

There is, however, another kind of pseudorobustness that has not often been discussed. This kind of pseudorobustness cannot be avoided by ensuring ontic independence between the kinds of evidence. And unlike the first kind of pseudorobustness, in every case of this kind of pseudorobustness the various ontically independent kinds of evidence that all individually support a hypothesis, when considered together, support a competitor of that hypothesis. We now turn to this troubling form of pseudorobustness.

4. Dyssynergystic Evidence. The philosophical literature on robustness has focused on ensuring that departures from the ideal robustness argument as described in section 3 and represented in figure 3 are avoided. The nearly universal idea seems to be that if the modes of evidence are OI (and the independent lines of evidence are confirming), then a robustness argument is warranted.

However, there are other ways that the structure of dependencies can depart from the ideal of figure 2, such that the validation of robustness provided above does not apply. The validation of robustness in section 2 simply established that conditional independence of $E1$ and $E2$ given H is sufficient for

https://doi.org/10.1086/692141 Published online by Cambridge University Press

robustness, not that it is necessary. It might be the case that even if this independence assumption is not satisfied, there is some other reason for thinking that more kinds of evidence increase confirmation, provided there are no common background assumptions. In what follows, however, we show that for a variety of simple departures from the ideal robustness scenario, it is possible to construct examples in which multiple confirmatory pieces of evidence are *dyssynergystic*: the conjoined evidence is less confirmatory than either of the conjuncts. In fact, in our examples, the conjoined evidence will be shown to be *disconfirming*, even though the individual conjuncts are confirmatory.

Diverse kinds of evidence that all individually support a hypothesis are *dyssynergystic* when the conjunction of all the evidence is less confirming than the conjunction of any proper subset of the evidence. The following is a formal definition.

> **Dyssynergystic Evidence (DE).** Two pieces of evidence, $E1$ and $E2$, are dyssynergystic if and only if $E1$ and $E2$ are ontically independent, both $E1$ and $E2$ confirm H, and $c(H, E1 \,\&\, E2) < c(H, E1)$ or $c(H, E1 \,\&\, E2) < c(H, E2)$, for a given confirmation measure c.

When OI evidence is DE, a robustness argument based on the OI evidence cannot be made. In what follows we provide three hypothetical cases to illustrate DE.

4.1. La Jolla Murder Mystery. The structure of dependencies represented in figure 4 is another form of departure from the ideal (on this graph and all following we have dropped the nodes representing auxiliary assumptions for the sake of simplicity).

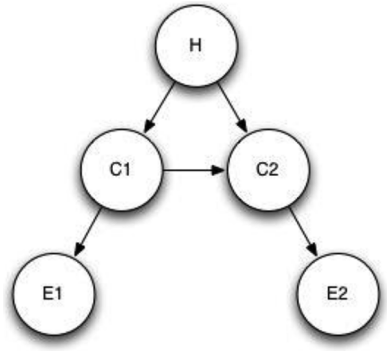


Figure 4. H no longer d-separates E1 and E2 because there is a dependency between C1 and C2 (and thereby between E1 and E2), even if we conditionalize on H.

https://doi.org/10.1086/692141 Published online by Cambridge University Press

TABLE 1. Likelihoods of Evidence in La Jolla Murder Mystery

	Pr(—|La Jolla)			Pr(—|~La Jolla)	
	Type A			Type A	
Blond	+	−	Blond	+	−
+	.2	.4	+	.4	0
−	.4	0	−	0	.6

The following case illustrates the dependencies represented in figure 4. A detective investigating a murder has an initial list of suspects, 50% of whom are from La Jolla. The detective acquires two pieces of evidence at the crime scene. A blond hair is found on the victim's body that must be the murderer's and a bit of the murderer's blood, determined to be type A. These pieces of evidence are OI. There are no substantive background assumptions in common between the tools used to analyze the blood and the tools used to analyze the hair.

The detective studies the suspects and discovers the following:

1. 60% of the suspects from La Jolla have blond hair.
2. 40% of the suspects from outside La Jolla have blond hair.
3. 60% of the suspects from La Jolla have type A blood.
4. 40% of the suspects from outside La Jolla have type A blood.
5. 20% of the suspects from La Jolla have both blond hair and type A blood.
6. 40% of the suspects from outside La Jolla have both blond hair and type A blood.

These conditional probabilities are summarized in table 1.

Let H be the hypothesis that the murderer is from La Jolla, $C1$ be that the murderer has blonde hair, $C2$ be that the murderer has blood type A, and let $E1$ be the blond hair and $E2$ be the type A blood. It is clear from the data that $\Pr(E1|H) > \Pr(E1|{\sim}H)$ and $\Pr(E2|H) > \Pr(E2|{\sim}H)$. It follows that $\Pr(H|E1) > \Pr(H)$ and $\Pr(H|E2) > \Pr(H)$. Each piece of evidence individually confirms H. It is also clear from the data that $\Pr(E1 \,\&\, E2|H) < \Pr(E1 \,\&\, E2|{\sim}H)$, which means $\Pr(H|E1 \,\&\, E2) < \Pr(H)$. The combined evidence is disconfirming. This, then, is a case of DE.[12] This case illustrates one possible way in which dyssynergy can arise.

12. At least, under the assumption that the measure of confirmation $c(H, E)$ will be an increasing function of the conditional probability $\Pr(H|E)$—a modest assumption for any viable probabilistic theory of confirmation. One might wonder why there is a directed dependency between C1 and C2. In sec. 2 we noted that directed dependencies arise as a result of causal relations. So, in this example, there is some feature that generates such a dependency: perhaps the murderer is part of a gang in La Jolla that selects members on the basis of both hair color and blood type.

https://doi.org/10.1086/692141 Published online by Cambridge University Press

4.2. Interacting Drugs. The structure of dependencies represented in figure 5 is another form of departure from the ideal robustness argument. The following case illustrates the dependencies represented in figure 5. Suppose you are a hospital physician, pondering a hypothesis about a patient's survival (let H be "the patient will live"), $C1$ is "the patient took drug X," $C2$ is "the patient took drug Y," and you have available two modes of evidence: the report from a nurse informing you that the patient is receiving drug X (call this $E1$) and the report from a life support machine informing you that the patient is receiving drug Y (call this $E2$). Both drug X and drug Y are known to help such patients, and so $\Pr(H|E1) > \Pr(H)$ and $\Pr(H|E2) > \Pr(H)$. However, as a matter of fact, drug X binds to drug Y, creating a lethal toxin that causes severe brain damage. Therefore, although $\Pr(H|E1) > \Pr(H)$ and $\Pr(H|E2) > \Pr(H)$, $\Pr(H|E1 \,\&\, E2) < \Pr(H)$. The evidence from the two modes is thus dyssynergystic. Hypotheses $C1$ and $C2$ (and also $E1$ and $E2$) are unconditionally independent, but they become dependent when one conditionalizes on H. Hypothesis H is a collider and thus does not d-separate E1 and E2.

One might respond to this case by noting that, precisely because the drugs interact in the way that we have described, the two modes of evidence are not in fact ontically independent. As we noted in section 1, OI requires modes to rely on "unrelated chunks of physics," to use Hacking's phrase, or different "theoretical presuppositions," to use Culp's phrase. The two modes of evidence in this case are OI, generally. But—goes this objection—the ontic independence of the two modes breaks down in this case, because of the way the drugs interact. (If so, then the case is not an example of dyssynergystic evidence.) This response entails that determining whether or not modes of

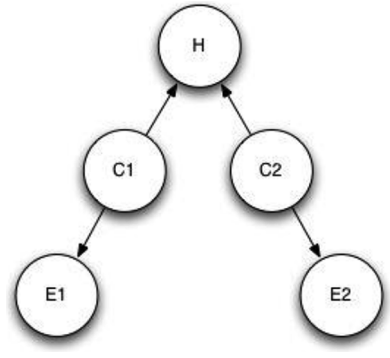


Figure 5. H does not d-separate E1 and E2 because, even though it lies on the only trail between E1 and E2, it is a collider.

https://doi.org/10.1086/692141 Published online by Cambridge University Press

evidence are OI cannot be assessed in a general manner but must take into account causal details of particular cases. (One could respond to the La Jolla murder mystery example in a similar way.) This way of understanding independence between modes of evidence departs significantly from the usual ways that OI is articulated, as described in section 1. CPI relativizes independence between modes of evidence to the hypothesis under investigation, whereas OI (as usually described) is not hypothesis relative in that sense. Perhaps it should be. In any case, such patches to the concept of OI amount to rendering it more like CPI (see sec. 4.4).

The two cases above are, admittedly, contrived. One might wonder to what extent such structures can be found in science. Our aim is not to suggest that such examples are especially prevalent. Rather, we simply wish to illustrate various possible ways that OI evidence can be dyssynergystic.

4.3. Hypothesis Competition. The structure of dependencies represented in figure 6 is another form of departure from the ideal robustness argument. Any straightforward case of hypothesis competition or theoretical change could satisfy the structure of figure 6.

4.4. Conditional Probabilistic Independence. Just as there is a way to guarantee the avoidance of pseudorobustness based on failure of OI (sec. 3), there is a way to guarantee the avoidance of pseudorobustness based on DE. A condition that rules out DE will be one that ensures that $\Pr(H|E1 \,\&\, E2) > \Pr(H|E1)$ and, mutatis mutandis, for $E2$. The graph of the ideal robustness argument in figure 2 represents just such a condition. We can formulate this condition in different terms as follows.

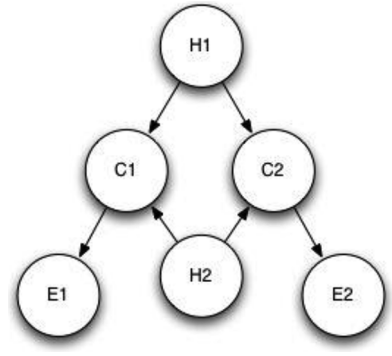


Figure 6. H1 does not d-separate E1 and E2 because of the presence of the alternate hypothesis H2 (and therefore an alternate trail between E1 and E2).

https://doi.org/10.1086/692141 Published online by Cambridge University Press

Conditional Probabilistic Independence (CPI). Two pieces of evidence, $E1$ & $E2$, are CPI iff these four conditions hold:

(i) $\Pr(H|E1) > \Pr(H)$,
(ii) $\Pr(H|E2) > \Pr(H)$,
(iii) $\Pr(E1 \,\&\, E2|H) = \Pr(E1|H) \times \Pr(E2|H)$,
(iv) $\Pr(E1 \,\&\, E2|{\sim}H) = \Pr(E1|{\sim}H) \times \Pr(E2|{\sim}H)$.

If $E1$ and $E2$ are CPI, then their conjunction must be more confirmatory than each individual conjunct. We provide a general proof of this in the appendix.

5. Discussion. We have shown that CPI provides a minimal justification to robustness arguments based on the concordance of diverse evidence. CPI guarantees the avoidance of DE. If evidence is CPI, then the second form of pseudorobustness discussed above is guaranteed to be avoided. However, note how modest are the gains of CPI and how demanding are its requirements. Although CPI is sufficient to avoid pseudorobustness, it does not warrant the epistemic prize that many have assumed robustness provides. Recall that those who rely on robustness arguments suppose that OI evidence that individually confirms a hypothesis gives a special epistemic oomph to a hypothesis when all of the evidence is considered together. That is what the no-miracles argument, or argument from coincidence, seems to suppose (sec. 1). CPI does not justify this special epistemic oomph. The argument here merely guarantees that CPI evidence provides some more support to a hypothesis than evidence that is not CPI.

Consider the various cases for realism based on robustness arguments, noted in section 1, including entity realism, causal realism, theory realism, and counters to the experimenter's regress and underdetermination worries. There is a wide gulf between

(i) Robustness arguments raise the probability of X and
(ii) Robustness arguments warrant realism about X.

One result of section 4 is the demonstration of i when the evidence is CPI. Another result of section 4 is the demonstration that ii cannot generally be the case when the evidence is merely OI. But even if the evidence is CPI, there is a large gulf between i and ii. Proving i lends no warrant to ii.

Thus our argument supports one aspect of a view recently proposed by Schupbach (2015, forthcoming), who argues that formal explications of robustness that rely on probabilistic independence are not compelling accounts of what is typically said to be gained by standard robustness arguments. About this we agree—satisfying the conditions of CPI does not guarantee that a hy-

https://doi.org/10.1086/692141 Published online by Cambridge University Press

pothesis gets the special epistemic oomph that is often said to result from robustness arguments—though we have shown that satisfying the conditions of CPI guarantees a boost in confirmation to the hypothesis in question.

However, we depart from Schupbach (forthcoming) in the following way. Schupbach appeals to a well-known problem that unconditional confirmational independence is not a good representation of robustness arguments, because multiple lines of evidence ($E1$ and $E2$) that confirm H and that are ontically independent nevertheless are unconditionally confirmationally dependent: for some measure of incremental confirmation c, $c(H, E1|E2) < c(H, E1)$. That is, $E1$ and $E2$ are unconditionally dependent on each other. Schupbach goes on to argue that condition iv above is insufficient as an explication of independence. His argument is compelling: even if H is false, there are many other possible hypotheses that could explain the converging evidence, and observing one of those lines of evidence ($E1$) increases the probability of observing the other ($E2$), conditional on some other hypothesis (H'). Nevertheless, CPI does not hold condition iv as sufficient for an increase in confirmation based on converging OI evidence, and it is not rendered sufficient by adding conditions i and ii. All of conditions i–iv are individually necessary subconditions of a sufficient condition, which is CPI. In short, contrary to Schupbach, we claim that probabilistic accounts of independent evidence such as CPI are perfectly good explications of one form of independent evidence, but with Schupbach we agree that this sort of independent evidence is a long way from accounting for the special epistemic oomph that many claim arises from robustness arguments.[13]

Moreover, the requirements of CPI are epistemically strong: in typical cases it is not clear that CPI is epistemically accessible to experimentalists. For CPI to be epistemically accessible, one must be able to determine if the probabilistic equalities and inequalities that constitute CPI are true. That is, one would have to know that $\Pr(E1\ \&\ E2|H) = \Pr(E1|H) \times \Pr(E2|H)$ and that $\Pr(E1\ \&\ E2|{\sim}H) = \Pr(E1|{\sim}H) \times \Pr(E2|{\sim}H)$, in addition to knowing that both $E1$ and $E2$ confirm H. Another way to put this is that one would have to know that the empirical scenario for which one has multiple lines of evidence has a structure of dependency relations that is like that modeled in figure 2 and does not have a structure of dependency relations like that of figures 3, 4, 5, and 6 (or other possible ways the CPI conditions can fail to hold). In rich empirical scenarios this is a demanding epistemic requirement.

13. Schupbach himself offers an alternative formal account of independent evidence in robustness arguments by framing robustness arguments as explanatory arguments and proceeding to offer a logic of such arguments based on recent formal work on explanation. Addressing his positive account would take us astray, but we direct the reader to Schupbach (forthcoming, sec. 3).

https://doi.org/10.1086/692141 Published online by Cambridge University Press

6. Conclusion. Two kinds of independence are appealed to in the literature on robustness: OI and CPI. There are two corresponding ways in which robustness arguments can fail. The first results from a failure of OI. The second is due to what we call DE. Both forms of pseudorobustness threaten the special epistemic oomph often thought to be gained by robustness. For each form of pseudorobustness, there is a way to avoid it by ensuring that certain conditions are satisfied.

We have shown that OI is insufficient for guaranteeing the special epistemic oomph often associated with robustness arguments. That is because of the possibility of DE. We have also shown that CPI is sufficient to warrant at least some increase in confirmation when diverse evidence is available. This increase in confirmation, though, is short of the special epistemic oomph often presupposed by robustness arguments.

CPI is a hypothesis-relative notion. Its epistemic significance comes from the fact that it rules out DE, which is again a hypothesis-relative notion. The notions of independence used by many philosophers in robustness-type arguments—versions of OI—are not hypothesis relative. The independence they rely on is based on a relationship between kinds of evidence, irrespective of the particular hypothesis being tested. A light microscope and an electron microscope are ontically independent sources of evidence *simpliciter*, not independent relative to some hypothesis.[14]

There is a tendency to take formal results akin to CPI as an explanation of the value of ontic independence in robustness arguments. For instance, Lloyd (2009) appeals to a probabilistic notion of independence similar to ours as an explanation of why it is valuable to test global climate models using OI evidence. More broadly, as we saw in section 1, most explications of robustness arguments are made in terms of OI; but since we have shown that OI is insufficient to rule out dyssynergystic evidence, it follows that OI is insufficient as a general basis for warranting claims about the special epistemic oomph often said to be gained by robustness arguments. A number of authors move from the ontic independence of kinds of confirming evidence in some scientific episode to a claim that the particular hypothesis under investigation is therefore more strongly confirmed; the unstated and not gen-

14. Howson and Urbach (1989) propose a probabilistic measure for the diversity of a set of evidence that is not hypothesis relative. The measure tracks the degree to which the pieces of evidence are unconditionally correlated. This measure is criticized in Wayne (1995), Myrvold (1996), and Fitelson (2001). If all the evidence is supposed to be evidence for a single hypothesis, then unconditional independence is an inappropriate condition. As Fitelson points out, newspaper reports and radio reports of the same baseball game are plausibly independent modes of evidence about the game, but they are not unconditionally independent. Myrvold patches the correlation view to emphasize independence conditional on a hypothesis, but this returns us to a hypothesis-relative conception of independence and, rightfully, cannot characterize ontic independence.

https://doi.org/10.1086/692141 Published online by Cambridge University Press

erally justified assumption is that confirmatory OI evidence provides the special epistemic oomph often associated with robustness arguments. OI does, at least, guarantee that the first form of pseudorobustness we discuss is avoided. But as long as dyssynergystic evidence is possible, the special epistemic oomph associated with robustness arguments is not warranted merely by ensuring OI, because of the second form of pseudorobustness. Some additional condition beyond OI must be appealed to—some condition that rules out dyssynergy (and, for most applications of robustness arguments, demonstrates that special epistemic oomph). CPI does, in fact, rule out dyssynergy but does not in addition warrant that special epistemic oomph. Moreover, CPI is a strong condition, and it is unclear how frequently experimentalists are in an epistemic position to know that CPI is satisfied.

Appendix

Proof

Assume that

$$\Pr(H|E1) > \Pr(H), \tag{A1}$$

$$\Pr(H|E2) > \Pr(H), \tag{A2}$$

$$\Pr(E1 \,\&\, E2|H) = \Pr(E1|H) \times \Pr(E2|H), \tag{A3}$$

and

$$\Pr(E1 \,\&\, E2|{\sim}H) = \Pr(E1|{\sim}H) \times \Pr(E2|{\sim}H). \tag{A4}$$

We want to show that $\Pr(H|E1 \,\&\, E2) > \Pr(H|E1)$. Assume for *reductio* that

$$\Pr(H|E1 \,\&\, E2) \leq \Pr(H|E1), \tag{A5}$$

$$\Pr(E1 \,\&\, E2|H) \times \Pr(H)/\Pr(E1 \,\&\, E2) \leq \Pr(E1|H) \times \Pr(H)/\Pr(E1) \tag{A6}$$

from (A5) and Bayes's theorem (BT);

$$\Pr(E1 \,\&\, E2|H)/\Pr(E1 \,\&\, E2) \leq \Pr(E1|H)/\Pr(E1) \tag{A7}$$

from (A6);

$$\Pr(E1 \,\&\, E2|H)/\Pr(E1|H) \leq \Pr(E1 \,\&\, E2)/\Pr(E1) \tag{A8}$$

https://doi.org/10.1086/692141 Published online by Cambridge University Press

from (A7);

$$\Pr(E2|H) \leq \Pr(E1 \ \& \ E2)/\Pr(E1) \tag{A9}$$

from (A3) and (A8);

$$\Pr(E2|H) \times \Pr(H)/\Pr(E2) > \Pr(H) \tag{A10}$$

from (A2) and BT;

$$\Pr(E2|H) > \Pr(E2) \tag{A11}$$

from (A10);

$$\Pr(E1 \ \& \ E2)/\Pr(E1) > \Pr(E2) \tag{A12}$$

from (A9) and (A11);

$$1 - \Pr(H|E1 \ \& \ E2) \geq 1 - \Pr(H|E1) \tag{A13}$$

from (A5);

$$\Pr(\sim H|E1 \ \& \ E2) \geq \Pr(\sim H|E1) \tag{A14}$$

from (A13);

$$\begin{aligned} &\Pr(E1 \ \& \ E2|\sim H) \times \Pr(\sim H)/\Pr(E1 \ \& \ E2) \\ &\geq \Pr(E1|\sim H) \times \Pr(\sim H)/\Pr(E1) \end{aligned} \tag{A15}$$

from (A14) and BT;

$$\Pr(E1 \ \& \ E2|\sim H)/\Pr(E1 \ \& \ E2) \geq \Pr(E1|\sim H)/\Pr(E1) \tag{A16}$$

from (A15);

$$\Pr(E1 \ \& \ E2|\sim H)/\Pr(E1|\sim H) \geq \Pr(E1 \ \& \ E2)/\Pr(E1) \tag{A17}$$

from (A16);

$$\Pr(E2|\sim H) \geq \Pr(E1 \ \& \ E2)/\Pr(E1) \tag{A18}$$

from (A4) and (A17);

$$\Pr(E2|\sim H) > \Pr(E2) \tag{A19}$$

from (A12) and (A18);

$$\Pr(\sim H|E2) \times \Pr(E2)/\Pr(\sim H) > \Pr(E2) \tag{A20}$$

https://doi.org/10.1086/692141 Published online by Cambridge University Press

from (A19) and BT;

$$\Pr(\sim H|E2) > \Pr(\sim H) \tag{A21}$$

from (A20);

$$1 - \Pr(H|E2) > 1 - \Pr(H) \tag{A22}$$

from (A21); and

$$\Pr(H|E2) < \Pr(H) \tag{A23}$$

from (A22).

There is a contradiction between (A2) and (A23), so we infer the falsity of (A5). Our four assumptions commit us to the claim that $\Pr(H|E1 \,\&\, E2) > \Pr(H|E1)$. A similar argument could easily be constructed to show that the assumptions also commit us to the claim that $\Pr(H|E1 \,\&\, E2) > \Pr(H|E2)$.

REFERENCES

Avery, O. T., C. M. MacLeod, and M. McCarty. 1944. "Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types: Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type III." *Journal of Experimental Medicine* 79:137–58.

Bechtel, William. 2002. "Aligning Multiple Research Techniques in Cognitive Neuroscience: Why Is It Important?" *Philosophy of Science* 69 (Proceedings): S48–S58.

Bovens, Luc, and Stephan Hartmann. 2004. *Bayesian Epistemology*. Oxford: Oxford University Press.

Cartwright, Nancy. 1983. *How the Laws of Physics Lie.* Oxford: Clarendon.

———. 1991. "Replicability, Reproducibility, and Robustness: Comments on Collins." *History of Political Economy* 23:143–55.

Claveau, Francois. 2013. "The Independence Condition in the Variety-of-Evidence Thesis." *Philosophy of Science* 80:94–118.

Culp, Sylvia. 1995. "Objectivity in Experimental Inquiry: Breaking Data-Technique Circles." *Philosophy of Science* 62:438–58.

Fitelson, Branden. 2001. "A Bayesian Account of Independent Evidence with Applications." *Philosophy of Science* 68 (Proceedings): S123–S140.

Fitz-James, P. 1960. "Participation of the Cytoplasmic Membrane in the Growth and Spore Formation of Bacilli." *Journal of Biophysical and Biochemical Cytology* 8:507–28.

Hacking, Ian. 1983. *Representing and Intervening.* Cambridge: Cambridge University Press.

Howson, Colin, and Peter Urbach. 1989. *Scientific Reasoning: The Bayesian Approach.* LaSalle, IL: Open Court.

Hudson, Robert G. 1999. "Mesosomes: A Study in the Nature of Experimental Reasoning." *Philosophy of Science* 66:289–309.

———. 2014. *Seeing Things*. Oxford: Oxford University Press.

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

Kuorikoski, Jaakko, and Caterina Marchionni. 2016. "Evidential Diversity and the Triangulation of Phenomena." *Philosophy of Science* 83:227–47.

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

https://doi.org/10.1086/692141 Published online by Cambridge University Press

Lloyd, Elisabeth. 2009. "Varieties of Support and Confirmation of Climate Models." *Aristotelian Society Supplementary Volume* 83:213–32.
Myrvold, Wayne. 1996. "Bayesianism and Diverse Evidence: A Reply to Andrew Wayne." *Philosophy of Science* 63:661–65.
Odenbough, Jay, and Anna Alexandrova. 2011. "Buyer Beware: Robustness Analyses in Economics and Biology." *Biology and Philosophy* 26:757–71.
Orzack, Steve, and Elliott Sober. 1993. "A Critical Assessment of Levin's *The Strategy of Model Building in Population Biology* (1966)." *Quarterly Review of Biology* 68:533–46.
Parker, Wendy. 2011. "When Climate Models Agree: The Significance of Robust Model Predictions." *Philosophy of Science* 78:579–600.
Pearl, Judea. 2000. *Causality: Models, Reasoning, and Inference.* Cambridge: Cambridge University Press.
Rasmussen, Nicolas. 1993. "Facts, Artifacts, and Mesosomes: Practicing Epistemology with the Electron Microscope." *Studies in History and Philosophy of Science* 24:221–65.
Salmon, Wesley. 1984. *Scientific Explanation and the Causal Structure of the World.* Princeton, NJ: Princeton University Press.
———. 1997. *Causality and Explanation.* Oxford: Oxford University Press.
Schupbach, Jonah. 2015. "Robustness, Diversity of Evidence, and Probabilistic Independence." In *Recent Developments in the Philosophy of Science: ESPA13 Helsinki*, ed. Uskali Mäki, Ioannis Votsis, Stéphanie Ruphy, and Gerhard Schurz, 305–16. Dordrecht: Springer.
———. Forthcoming. "Robustness Analysis as Explanatory Reasoning." *British Journal for the Philosophy of Science*.
Snyder, Laura. 2005. "Consilience, Confirmation, and Realism." In *Scientific Evidence: Philosophical Theories and Applications*, ed. Peter Achinstein. Baltimore: Johns Hopkins University Press.
Sober, Elliott. 2008. *Evidence and Evolution*. Cambridge: Cambridge University Press.
Stegenga, Jacob. 2011. "The Chemical Characterization of the Gene: Vicissitudes of Evidential Assessment." *History and Philosophy of the Life Sciences* 33:105–27.
Strawson, Peter. 1952. *Introduction to Logical Theory*. London: Methuen.
Wayne, Andrew. 1995. "Bayesianism and Diverse Evidence." *Philosophy of Science* 62:111–21.
Weisberg, Michael. 2006. "Robustness Analysis." *Philosophy of Science* 73:730–42.
Whewell, William. 1857. *The History of the Inductive Sciences, from the Earliest to the Present Time*. New York: Appleton.
Wimsatt, William. 1981. "Robustness, Reliability, and Overdetermination." In *Scientific Inquiry and the Social Sciences*, ed. Marilynn B. Brewer and Barry E. Collins. San Francisco: Jossey-Bass.
Woodward, James. 2006. "Some Varieties of Robustness." *Journal of Economic Methodology* 13:219–40.

https://doi.org/10.1086/692141 Published online by Cambridge University Press