Robustness, Diversity of Evidence, and Probabilistic Independence

Jonah N. Schupbach

Abstract In robustness analysis, hypotheses are supported to the extent that a result proves robust, and a result is robust to the extent that we detect it in diverse ways. But what precise sense of *diversity* is at work here? In this paper, I show that the formal explications of evidential diversity most often appealed to in work on robustness – which all draw in one way or another on probabilistic independence – fail to shed light on the notion of diversity relevant to robustness analysis. I close by briefly outlining a promising alternative approach inspired by Horwich's (1982) *eliminative* account of evidential diversity.

1 Robustness Analysis in Science

To verify that results are not simply artifacts of the particular means used to detect them, scientists often attempt to duplicate those results using other, diverse means. To the extent that a result is detected via numerous, diverse means, it is said to be *robust*. *Robustness analysis* (henceforth, "RA") is a mode of reasoning in which one supports a hypothesis via an analysis of the conditions under which a result is robust.

Examples of RA from scientific practice abound. Famously, biologist Richard Levins (1966) proposes RA as a general means for deciphering, when using simplified models to study complex systems, whether a result "depends on the essentials of a model or on the details of the simplifying assumptions:"

> [W]e attempt to treat the same problem with several alternative models each with different simplifications but with a common biological assumption. Then, if these models, despite their different assumptions, lead to similar results we have what we can call a robust theorem which is relatively free of the details of the model. Hence our truth is the intersection of independent lies.

Jonah N. Schupbach

Department of Philosophy, University of Utah, Salt Lake City, Utah 84112 USA, e-mail: jonah.n.schupbach@utah.edu

As a specific example of this use of RA in modeling, Weisberg and Reisman (2008) discuss work in support of the Volterra Principle:

Volterra Principle   *Ceteris paribus*, if a two-species, predator-prey system is negatively coupled, then a general biocide will increase the abundance of the prey and decrease the abundance of predators.

The earliest predator-prey models demonstrating the Volterra Principle included several unrealistic, simplifying assumptions – including single growth- and death-rates for prey and predators respectively as well as linear functions relating prey capture-rates to number of predators and predator birth-rates to number of captures. As Weisberg and Reisman emphasize, however, the Volterra Principle can be demonstrated robustly across more complicated models that do away with various of the simplifying assumptions. Some of these models, for example, "add in terms representing predator satiation, the ability of prey to seek cover, multiple sources of food for the predator, or even complex adaptive behaviors such as learning" (116).

In this example, the robust result is the observed qualitative behavior of various models – commonly interpreted as the increased relative size of prey population when a general biocide is introduced. The diverse means which detect this result are the mathematical models themselves. And the hypothesis most apparently supported by the fact that this result is robust is the Volterra Principle – the biological claim, not to be confused with claims about the mathematical representations of biological systems.

The case of Brownian motion provides a rather different example of RA in science – also commonly discussed by philosophers of science. When suspended in a fluid medium, sufficiently small particles display continual and seemingly random movements. Upon discovering this phenomenon, the botanist Robert Brown surmised that the motions were due to the particular (uniquely shaped) pollen granules he was observing. However, he found that the movements persisted across experiments using other particles – first using other types of pollen, then other organic materials, and eventually using inorganic particles. Over the next 75 years, other experimenters showed that the "Brownian motion" was also robust over changes in the fluid medium, container used, means of suspending the particles, environmental conditions around the container, and so on. Eventually, scientists working in the wake of Einstein's *annus mirabilis* appealed to the robustness of the Brownian motion in order to support the idea that there were deeper, unobservable agitations of molecules within the medium – just as the evident rocking of a far off ship betrays imperceptibly distant waves on the sea (Perrin 1913, 83).

In this case, the Brownian motion is the result detected robustly. The diverse means of detection are the various experiments used; regarding these experiments, the Brownian motion is notably robust across certain changes to the experimental apparatus (type of particle, medium, container, lighting, etc.) and sensitive to others (size of particle, temperature of medium). And it is upon analyzing these conditions of robustness that Perrin (1913, 86) says we are "forced to conclude," consonant with Einstein's molecular explanation, that there are internal, unobservable movements in the medium.

Many other example RAs include cases from cognitive psychology (Stolarz-Fantino et al. 2003; Crupi et al. 2008), "arguments from coincidence" in physics (Hacking 1983; Cartwright 1991; Mayo 1996), experimental biology (Culp 1994), climate science (Lloyd 2010; Parker 2011), and modeling in economics (Woodward 2006; Kuorikoski et al. 2010). As should be evident from the range of these cases, I mean for the terms "results" and "means of detection" to be quite generic. The results in question could be observations, measurements, predictions, theorems, and so on. Correspondingly, the means of detecting such results could include experiments, laboratory instruments, sensory modalities, derivations (from axioms, models, theories, etc.), axiomatic systems, computer simulations, and formal models amongst other things.

2 Evidential Diversity and RA-Diversity

The intuition motivating the use of RA is that we can gain confirmation through diversity; certain hypotheses (e.g., Einstein's molecular explanation of Brownian motion) are supported to the extent that a result proves robust, and results are robust to the extent that we detect them in diverse ways. But what precise sense of diversity is involved in RAs? Philosophers have offered many distinct accounts of evidential diversity. And many of these accounts plausibly capture legitimate senses in which we speak of evidence as being diverse. But is there a single sense of evidential diversity that drives our reasoning in RAs? For the sake of this paper, we will work on the optimistic assumption that there is.

The hope is that a precise account of such "RA-diversity" would illuminate the normative import of actual RAs. But in order to stand any chance of doing so, such an account must be held accountable to scientific practice. A particular account of evidential diversity may, for example, specify precise conditions under which diverse bodies of evidence provide strong confirmation for relevant hypotheses. But such an account will clearly not illuminate RA if it relies on a notion of diversity that does not fit with actual cases of RA in science. Such an account may shed light on the confirmational import of certain diverse bodies of evidence, just not on RA-diverse bodies of evidence.

This main section of the present paper applies this consideration in criticizing the most common formal accounts of RA-diversity. These formalize diversity using probabilistically-precise notions of independence. Moreover, several of these accounts imply interesting senses in which diverse bodies of evidence may be specially confirmatory. The problem is that these accounts fail to capture paradigmatic cases of RA from science. To show this in each case, I will return to the above examples of Brownian motion and the Volterra Principle.

2.1 Unconditional Probabilistic Independence

As a first attempt at explicating RA-diversity, we might take a cue from Levins's quote and surmise that some precise notion of "independence" is at work. Most simply, one might say that if two means of detection are RA-diverse in the relevant sense, then they are (unconditionally) probabilistically independent of one another.

To make this idea more precise, let R be a proposition describing the result that has been robustly detected by various means. Then, let us denote the proposition that this result is detected using the k'th means of detection as R_k. According to the *unconditional probabilistic independence* account, if two means of detection are RA-diverse, then the fact that R is detected via means i should have no bearing whatever on the probability that R will be detected using means j: $Pr(R_i \& R_j) = Pr(R_i) \times Pr(R_j)$ – which (assuming that $Pr(R_i), Pr(R_j) > 0$) entails that $Pr(R_i) = Pr(R_i|R_j)$ and $Pr(R_j) = Pr(R_j|R_i)$.[1]

In their critique of Levins's discussion of RA, Orzack and Sober (1993, 539-40) consider and quickly dismiss this explication; they argue that, by requiring the various models to share "a common biological assumption," Levins's "'Protocol' for the discovery of robust predictions guarantees that the models under consideration are *not* independent."[2] In a bit more detail and put more explicitly in Bayesian terms, when such a model implies a particular result in RA settings, we consider it possible (and perhaps even plausible) that the result is driven by the essential core of the model – i.e., Levins's common biological assumption. But then whether or not we get a result from one of the models will manifestly provide relevant information with regards to whether we will get the result from another model with the same common core. Typically, the fact that we have detected R with one such model will increase the probability that we will detect R using another: $Pr(R_i) < Pr(R_i|R_j)$. Importantly, this may be true despite the fact that these models are considered diverse for the sake of RA.

For similar reasons, RA-diverse experiments in the case of Brownian motion also fail to be unconditionally independent. Take any two of these experiments, say those suspending dust particles in water and those suspending them in ethanol. Although these are diverse in the relevant sense that makes it appropriate for scientists, like Perrin, to cite them as part of their RA, the respective results of these experiments may strongly inform one another. This will be the case, in fact, so long as one allows that other factors (besides whether to use water or ethanol as the medium) may potentially influence whether one observes the result. These experiments actually share in common the vast majority of their respective traits – type of particle used, means of suspending the particle, lighting conditions, etc. – which may all be

---

[1] For clarity and ease of exposition, I leave the background beliefs term implicit in all Bayesian formulations.

[2] Orzack and Sober also criticize an alternative explication according to which two models are diverse only if they are *logically* independent. The fact that RA-diverse models may involve contrary simplifying assumptions spells trouble for this account; e.g., "A model with the assumption of random mating is not logically independent of a model with the assumption that mating is assortative; the reason is that the truth of one entails the falsity of the other."

seen as potentially relevant in affecting the result. But then, observations of Brownian movements in water may greatly increase the probability that one will observe Brownian movements in ethanol. In this case again, perfectly RA-diverse means of detection may be such that $Pr(R_i) < Pr(R_i|R_j)$.

2.2 Reliability Independence

While this initial effort thus fails, there are subtler ways one can attempt to use probabilistic independence to explicate RA-diversity. Wimsatt (1994, 197) offers such an account, proposing "that the *probability of failure* of the different means of access should be independent." This account arguably doubles as a more accurate interpretation of Levins's thought that RA requires "independent *lies*." The lies, the ways that each means of detection could lead us astray, are the things that are required to be independent between RA-diverse means of detection.[3]

This *reliability independence* account is importantly distinct from the unconditional independence account above. Instead of enforcing the stringent condition that the results of the various means of detection be entirely irrelevant to one another, this account just requires that if the means in question lead us astray, they do so for independent reasons. More precisely, learning that one of our means of detection has misled us has no effect on the probability that the other means of detection will mislead us. Each means of detection is or isn't reliable, independent of the others.

One nice feature of this account is the straightforward way in which it reveals the epistemic appeal of diversity. The justification that a hypothesis receives from evidence that is diverse in this sense has all the logical advantage of webs over chains. Whereas a linear chain of justification can be no stronger than its weakest link, a web of independent lines of justification is no weaker than its strongest member. Wimsatt (1981, 49-50) offers a quick probabilistic demonstration of this as follows: Assume that we have n means, all of which detect a result. Now assume that these means are reliability independent. Naturally, these means are imperfect, and so each may lead us astray with some probability; for simplicity, assume that they each may lead us astray with the same probability p_0. Now, if the common result these means are all detecting is misleading, then all n means of detection are going astray. Because they do so independently of one another, we know that the probability of this happening is $p_p = p_0^n$. Wimsatt concludes, "But p_0 is presumably always less than 1; thus, for $n > 1$, p_p is always less than p_0. Adding alternatives for redundancy always increases reliability."

Unfortunately, while reliability independence manifestly explicates an important notion of evidential diversity, it too fails to capture the notion of RA-diversity. First, return to the example of Brownian motion, and again consider any two of the RA-diverse means of detection used by Brown; this time, let us compare experiments in

---

[3] Bovens and Hartmann (2003, 96-97) offer an in-depth formal exploration of this notion of evidential diversity, and Kuorikoski et al. (2010, 544-45) follow Wimsatt in adopting this account as an explication of RA-diversity.

which a variety of pollen granules were suspended in water with those in which a variety of inorganic materials were suspended in water. These experiments are again cited as diverse in the sense required for RA. Yet, their respective reliabilities surely have a bearing on one another. To be sure, they could be unreliable for different reasons. But there are any number of *common* reasons that they might be unreliable too; there are many possible confounding factors that could be driving the result in both cases. Both could be misleading us due to the way the particles are being suspended, due to the use of the same medium, due to the use of the same environmental conditions surrounding the apparatus, etc. To the extent that we are aware of such overlapping sources of potential unreliability, learning that one of these experiments is leading us astray provides relevant information when deciding whether to trust the other. In particular, such information will often greatly reduce our estimate of how reliable the other is.

The reliability independence account encounters the same problem when trying to model RA-diversity in examples from modeling. Two RA-diverse predator-prey models that demonstrate the Volterra Principle may differ only on whether they involve a particular simplifying assumption, say the assumption that prey capture-rate increases linearly with number of predators. A model that is more realistic in this one regard and the fully simplified model will share many potential sources of unreliability when it comes to modeling the complex predator-prey dynamics (e.g., not allowing prey to take cover or learn). But then learning that one of the models is unreliable will potentially greatly increase our confidence that the other is too. In general, fully RA-diverse means of detection can nonetheless be susceptible to many of the same potential confounds; in such cases, learning that one of our means of detection is unreliable will often greatly increase the likeliness that other of our means of detection is similarly unreliable.

2.3 Confirmational and Conditional Independence

Lloyd (2009; 2010) has recently proposed a third independence-based account worth considering here. She proposes that RA-diversity amounts to *confirmational independence*, as explicated by Fitelson (2001). This sense is defined relative to a particular hypothesis (call it H), which we may think of as the hypothesis intuitively supported via the RA. Two means of detection are RA-diverse, according to this account, only if their results incrementally confirm / disconfirm H (raise / lower H's probability) to the same extent regardless of whether we have detected and learned the results using the other means. More formally (using the notation we introduced in Section 2.1 above, and where c stands in for an adopted Bayesian measure of incremental confirmation): if the i'th and j'th means of detection are RA-diverse with respect to H, then $c(H, R_i | R_j) = c(H, R_i)$ and $c(H, R_i | R_j) = c(H, R_j)$.[4]

---

[4] Notation: $c(x, y)$ measures the degree of confirmation that y lends to x; $c(x, y|z)$ measures the degree of confirmation that y lends to x, conditional on (or given that) z.

As with Wimsatt's account of evidential diversity, this idea nicely illuminates the normative appeal of diversifying our evidence. Accepting any of the most defensible and popular Bayesian measures of confirmation as c, and assuming that each detection of R individually confirms H to some extent, one can prove that confirmationally independent means of detection jointly confirm H to a greater extent than either means of detection does individually: $c(H, R_i \& R_j) > c(H, R_i)$ and $c(H, R_i \& R_j) > c(H, R_j)$ (Fitelson 2001, S131).

Before evaluating this account, it is worth mentioning that confirmational independence has a direct connection to conditional probabilistic independence, relative to H. As Fitelson (2001, S129) clarifies, "screening-off by H of R_i from R_j is a sufficient condition for R_i and R_j to be mutually confirmationally independent regarding H."[5] By "screening-off," Fitelson has in mind the standard Reichenbachian (1956, 158-59) notion, implying the dual conditional independencies: $Pr(R_i \& R_j | H) = Pr(R_i | H) \times Pr(R_j | H)$ and $Pr(R_i \& R_j | \neg H) = Pr(R_i | \neg H) \times Pr(R_j | \neg H)$.

Unfortunately, confirmational independence also does not fit with the notion of RA-diversity. Consider again two experiments from the Brownian motion case. Let R_1 describe the fact that we have observed Brownian motion using the uniquely shaped and sized pollen granules of *Clarkia pulchella* (the wildflower first used by Brown in his experiments), and let R_2 be the proposition that we have witnessed the same motions using other types of pollen. While these two results are diverse in the sense that makes them crucial to establishing the robustness of Brownian motion (and both mentioned explicitly as such by Perrin), they are evidently not confirmationally independent regarding H: Perrin's inferred hypothesis that there are unobservable movements internal to fluid media. To assert that they are would be to claim that his hypothesis is supported to the same extent by R_1, regardless of whether we know R_2. But while H may be strongly supported by experiments observing the jostling of granules of a particular type of pollen, it plausibly does not gain nearly so much support from such an observation if one has already witnessed the jostling using several other types of pollen: $c(H, R_1 | R_2) < c(H, R_1)$. On the contrary, the more pollens that we have already observed in motion, the less a confirmatory impact on H future experiments using pollens will have.

The following observation helps us to see why this account does not work from another angle. The fact that these diverse means of detection are not confirmationally independent regarding H implies that their results also will not be screened-off by H. Here, we can pinpoint the feature of screening-off that generally will not be satisfied by these experiments, the clause that asserts that R_1 and R_2 should be independent conditional on $\neg H$: $Pr(R_1 \& R_2 | \neg H) = Pr(R_1 | \neg H) \times Pr(R_2 | \neg H)$. If H is false, there remain many potential reasons why we might see particles dance about in fluids. Take for example the idea H' that this motion is due to the nature of the suspended particle. Conditional on H', the observation of Brownian motion using various pollens will greatly increase the probability of witnessing it in other pollens: $Pr(R_1 | H') \ll Pr(R_1 | H' \& R_2)$. After all, on this hypothesis, this motion is attributable

---

[5] I have replaced Fitelson's notation with our own. It should be noted that Fitelson suggests this relation as a condition of adequacy on measures of confirmation, as opposed to proving and presenting it as a theorem that follows robustly (!) for all candidate measures.

to some aspect of the suspended particle; but then witnessing it across samples of pollen will make us more confident that all pollens share the relevant attribute (e.g., the sexual drive or vital force inherent in the particles). More generally, given that H is false, we might still observe R according to several alternative possibilities. And RA-diverse means of detecting R can be probabilistically relevant to one another conditional on these other possibilities.

Similar points weigh against the idea that confirmational or conditional independence explicates RA-diversity in cases from modeling. For example, conditional on the Volterra Principle being false, there could be several reasons why our models are displaying qualitative behavior interpreted in accordance with this principle. For example, perhaps this behavior is in part an artifact of the unrealistic assumption that prey are borne at a single constant rate. Conditional on the hypothesis that this partially drives our result, two RA-diverse models that both assume single growth-rates for prey (e.g., two models differing only on whether they represent predator satiation) may be substantially probabilistically relevant to one another; if one provides the result, this may greatly increase the probability that the other will too.

2.4 Partial Independence?

One might think that the problem is just that we have framed the above accounts as requiring *full* unconditional, reliability, or confirmational independence. But perhaps we can make these accounts more defensible by adjusting them to measure degrees of RA-diversity. Two means of detection are RA-diverse, we might say, *to the extent* that they approach full unconditional, reliability, or confirmational independence. Wimsatt (1981, 46) may have just this sort of move in mind when he writes, "All these procedures require at least *partial independence* of the various processes across which invariance is shown."

But note that the reasons above for why these accounts fail have little to do with the fact that we require full independence. The problem, in other words, is not that the RA-diverse means of detection in these paradigmatic cases fall just short of full independence in one of the three senses. In fact, we have seen that means of detection that are recognizably and clearly RA-diverse may not even come remotely close to being independent in any of the above three senses. Nor is it at all clear that we would end up with means of detection that are more RA-diverse if we sought those that came closer to full unconditional, reliability, or confirmational independence. In fact, in all of the examples proffered, the means of detection are intuitively fully diverse in the sense required for them to do their work in RA. When Perrin cites experiments detecting Brownian motion using organic particles, and then those using inorganic particles, there is a sense in which these means of detection are perfectly diverse in the sense needed for these to have their respective roles in Perrin's larger RA. And there is no clear reason to think that Perrin's cited means could have been improved in their RA-diversity roles had they been less dependent in one of the above probabilistically precise senses.

What is this general role that means of detection are meant to perform by virtue of their RA-diversity? The answer that I want to explore is, in a word, *elimination*. While the experiments that Perrin cites and the models used to demonstrate the Volterra Principle are actually, in several cases, overall quite similar to one another, they inevitably remain distinct in ways that make them useful for ruling out *H*'s potential competitors. In fact, there is already an account of evidential diversity that fits well with this eliminative idea. The following, closing section briefly explores this account and whether it holds more promise as an explication of RA-diversity.

3 Toward an Alternative Explication of RA-Diversity

Horwich (1982, 118-22) proposes an account of evidential diversity. While this account is probabilistic, it does not make use of probabilistic independence. The central notion in Horwich's account is instead that of *elimination*; diverse bodies of evidence, according to Horwich, "tend to eliminate from consideration many of the initially most plausible, competing hypotheses" (118). Probabilistically, Horwich represents "initially plausible" competing hypotheses as those with substantial prior probabilities, and he identifies diverse evidence with evidence that takes low likelihoods conditional on competing hypotheses.

Horwich argues for the normative appeal of diverse evidence using his account in the following way. Let E_D describe a more *eliminatively diverse* body of evidence than E_N relative to our favored target hypothesis H_1 and its competitors $H_2, H_3, ..., H_k$. We can compare the probabilistic effects of both sets of evidence on H_1 by comparing $Pr(H_1|E_D)$ to $Pr(H_1|E_N)$. Horwich stipulates that the alternative hypotheses form a partition $\{H_1, H_2, ..., H_k\}$ and that H_1 implies E_N and E_D, so that $Pr(E_N|H_1) = Pr(E_D|H_1) = 1$. Under these conditions, we have:

$$\frac{Pr(H_1|E_D)}{Pr(H_1|E_N)} = \frac{Pr(H_1)}{Pr(H_1)} \times \frac{Pr(E_D|H_1)}{Pr(E_N|H_1)} \times \frac{Pr(E_N)}{Pr(E_D)} = \frac{Pr(E_N)}{Pr(E_D)}$$

$$= \frac{Pr(H_1) + Pr(H_2)Pr(E_N|H_2) + ... + Pr(H_k)Pr(E_N|H_k)}{Pr(H_1) + Pr(H_2)Pr(E_D|H_2) + ... + Pr(H_k)Pr(E_D|H_k)}.$$

Comparing like terms between the numerator and denominator of this ratio, the only terms that may affect a difference between $Pr(H_1|E_D)$ and $Pr(H_1|E_N)$ are the likelihoods relative to H_1's competitors. Consequently, to the extent that all of those hypotheses with considerable values of $Pr(H_i)$ are such that $Pr(E_D|H_i) < Pr(E_N|H_i)$, it will tend to be the case that $Pr(H_1|E_D) > Pr(H_1|E_N)$. But that is just to say that the more eliminatively diverse the evidence in this case, the more confirmation it will tend to bestow upon H_1.

If we use this account to explicate RA-diversity, we have that means of detection (cited in a RA to *H*) are diverse insofar as they are able to rule out *H*'s competitors. On such an account, it is not so important that means of detection are *strongly* diverse or sufficiently distinct in some sense separated from considered hypothe-

ses. What really matters for RA-diversity is that the means (which may actually be quite similar in most respects) are different in just the sense required to rule out *H*'s salient competitors. Accordingly, when seeking to increase the RA-diversity of our evidence, we search for a new way of detecting *R* that rules out some of *H*'s still-standing competitors.

Such an eliminative account of RA-diversity makes better sense of standard cases of RA in science. Many of the RA-diverse means of detecting Brownian motion are overall just not that diverse; indeed, these means may be identical in all respects other than some modest change – e.g., in the particle suspended or mode of suspending it. This is why accounts that require RA-diverse means to be strongly diverse (often in a sense that pays no attention to the relevant hypotheses) run quickly into trouble in this case. These means of detection are clearly eliminatively diverse, however. When Perrin cites experiments on *Clarkia pulchella*, and then increments the RA by citing experiments on other varieties of pollen, he is not doing so because these experiments are strongly heterogeneous overall, but because they are *relevantly* different than one another. The latter rules out a potential confounding hypothesis left standing by the first – viz., that the motion is attributable to the unique form of *Clarkia pulchella* granules.

Similarly, when seeking to confirm the Volterra Principle, RA-diverse models may be identical but for some modest difference. By utilizing these RA-diverse models, we rule out confounding hypotheses pertaining to our result left standing by any subset of the models used alone. Notably, we alleviate worries that our result is an artifact of a simplifying assumption common to some subset of our models by duplicating that result using a new model that does not share that assumption.

Though the eliminative account provides a prima facie more promising approach to explicating RA-diversity, much work remains for any satisfactory account in this vein. Perhaps most obviously, we would ultimately like a more subtle demonstration of the normative impact of eliminative diversity as it relates to RAs. There are several features of Horwich's demonstration that make it less informative regarding RAs. First, the general setting of Horwich's result has us comparing two bodies of evidence, one more and one less diverse, at the end of the collection process. But in practice, we are rarely at the end of an RA, and we are not directly interested in comparing two hypothetical bodies of evidence. The normative intuition that we would like to test is that *H* receives more confirmation with each increment of RA; consequently, a more informative account would examine the confirmatory effects on *H* of adding RA-diverse means of detecting the result to our working body of evidence.

Second, Horwich's demonstration hinges on some very specific problematic assumptions. In RAs, it is not obvious that we should require *H* to imply the detected results in question; this assumption will either need to be weakened or it will require further motivation. Nor is it obvious that all RAs involve hypotheses that compete in the sense of being mutually exclusive; in fact, in the Brownian motion case, Perrin's favored hypothesis is consistent with many of the competing hypotheses that get ruled out (e.g., it's possible that the motion is affected both by unobservable motions internal to the medium and by a vital force inherent in the particle used).

Horwich assumes too that the hypotheses before us exhaust the possibility space (they form a partition). But of course, in actual cases of RA, more often than not, this is not the case. And we would accordingly like an account that informs us of the epistemic import of RA-diversity in cases involving a catch-all hypothesis.

Third, Horwich's account tells us that, under all of the above conditions, more diverse evidence does indeed tend to bestow more confirmation on the relevant sort of hypothesis. But it would be nice if our account could tell us more than this. Is RA confirmatory under other conditions? And what determines the extent of confirmation that an increment of RA provides?

Finally, Fitelson (1996) suggests that Horwich's account is more properly viewed as an explication of the logical effects of diverse evidence than an explication of diverse evidence. There is an "intuitive notion" of evidential diversity that underlies and motivates Horwich's discussion of eliminative diversity. However, argues Fitelson, this intuitive notion proves elusive, and so Horwich's account is at best incomplete. A more appealing account of RA-diversity might provide the missing pieces here, starting from a more intuitive notion of diversity and then showing that something like Horwich's eliminative account captures in part the logical implications of such diversity. All of these considerations point to ways in which a fuller account of RA-diversity along the eliminativist lines will need to expand upon Horwich's account.

References

Bovens, L., & Hartmann, S. (2003). *Bayesian Epistemology*. Oxford: Oxford University Press.

Cartwright, N. (1991). Replicability, reproducibility, and robustness: Comments on Harry Collins. *History of Political Economy*, 23(1), 143–155.

Crupi, V., Fitelson, B., & Tentori, K. (2008). Probability, confirmation, and the conjunction fallacy. *Thinking and Reasoning*, 14(2), 182–199.

Culp, S. (1994). Defending robustness: The bacterial mesosome as a test case. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, 1: Contributed Papers, 46–57.

Fitelson, B. (1996). Wayne, Horwich, and evidential diversity. *Philosophy of Science*, 63(4), 652–660.

Fitelson, B. (2001). A Bayesian account of independent evidence with applications. *Philosophy of Science*, 68(3), S123–S140.

Hacking, I. (1983). *Representing and Intervening*. Cambridge: Cambridge University Press.

Horwich, P. (1982). *Probability and Evidence*. Cambridge: Cambridge University Press.

Kuorikoski, J., Lehtinen, A., & Marchionni, C. (2010). Economic modelling as robustness analysis. *British Journal for the Philosophy of Science*, 61(3), 541–567.

Levins, R. (1966). The strategy of model building in population biology. *American Scientist*, 54(4), 421–431.

Lloyd, E. A. (2009). Varieties of support and confirmation of climate models. *Proceedings of the Aristotelian Society, Supplementary Volumes*, 83, 213–232.

Lloyd, E. A. (2010). Confirmation and robustness of climate models. *Philosophy of Science*, 77(5), 971–984.

Mayo, D. G. (1996). *Error and the Growth of Experimental Knowledge*. Chicago: University of Chicago Press.

Orzack, S. H., & Sober, E. (1993). A critical assessment of Levins's *The Strategy of Model Building in Population Biology* (1966). *The Quarterly Review of Biology*, 68(4), 533–546.

Parker, W. S. (2011). When climate models agree: The significance of robust model predictions. *Philosophy of Science*, 78(4), 579–600.

Perrin, J. (1913). *Les Atomes* (trans: Hammick, D. Ll.). Woodbridge, Conn: Ox Bow Press.

Reichenbach, H. (1956). *The Direction of Time*. Berkeley, Cal: University of California Press.

Stolarz-Fantino, S., Fantino, E., Zizzo, D. J., & Wen, J. (2003). The conjunction effect: New evidence for robustness. *The American Journal of Psychology*, 116(1), 15–34.

Weisberg, M., & Reisman, K. (2008). The robust Volterra principle. *Philosophy of Science*, 75(1), 106–131.

Wimsatt, W. C. (1981). Robustness, reliability, and overdetermination. In M. B. Brewer and B. E. Collins (Eds.), *Scientific Inquiry and the Social Sciences* (pp. 125–163). New York: Jossey-Bass. Page references are to the version reprinted in (Wimsatt 2007).

Wimsatt, W. C. (1994). The ontology of complex systems: Levels of organization, perspectives, and causal thickets. *Canadian Journal of Philosophy*, 24(sup1), 207–274. Page references are to the version reprinted in (Wimsatt 2007).

Wimsatt, W. C. (2007). *Re-Engineering Philosophy for Limited Beings*. Cambridge, Mass: Harvard University Press.

Woodward, J. (2006). Some varieties of robustness. *Journal of Economic Methodology*, 13(2), 219–240.