Robustness Analysis as Explanatory Reasoning

Jonah N. Schupbach

Abstract

When scientists seek further confirmation of their results, they often attempt to duplicate the results using diverse means. To the extent that they are successful in doing so, their results are said to be *robust*. This paper investigates the logic of such "robustness analysis" [RA]. The most important and challenging question an account of RA can answer is what sense of evidential diversity is involved in RAs. I argue that prevailing formal explications of such diversity are unsatisfactory. I propose a unified, explanatory account of diversity in RAs. The resulting account is, I argue, truer to actual cases of RA in science; moreover, this account affords us a helpful, new foothold on the logic undergirding RAs.

1 Robustness Analysis in Science

In 1827, the botanist Robert Brown focused his microscope on a sample of pollen granules suspended in water. His aim was to get a better glimpse at the unique form of these particular granules, but his attention was instead drawn to their surprising motion, "consisting not only of a change of place in the fluid, manifested by alterations in their relative positions, but also not unfrequently of a change of form in the particle itself" ([1827], pp. 466-67). After confirming that the motions were not simply due to currents or evaporation of the water, Brown hypothesized that the motion was characteristic of the *Clarkia pulchella* pollen he was observing, this having an unusually large size and cylindrical shape. However, he quickly ruled out this idea when, placing the pollen of several other plants under the same conditions, Brown noted that "in all these plants

particles were found, [...] varied in form from oblong to spherical, having manifest motions similar to those already described." Through a series of other experiments, Brown tested various other potential explanations of the perceived motion (e.g., sexual drive inherent in pollen) and, continuing to observe the motion in each case, dismissed these ideas as well. For Brown, this process culminated in experiments detecting the same motion using various inorganic materials (including "a fragment of the Sphinx"!), thus disconfirming the idea that vital forces caused the motion. Brown concludes, "in every mineral which I could reduce to a powder, sufficiently fine to be temporarily suspended in water, I found these [moving] molecules more or less copiously" (p. 472).

In the following decades, physicists continued to explore the phenomenon of "Brownian motion". Additional experiments ruled out further hypotheses, revealing that the motion persisted in spite of changes to the container used, medium used, means of suspending the particles, lighting and general environmental conditions surrounding the experiment, etc. Circa 1900, this same motion had been detected in such a diverse variety of ways, proving robust across changes to a host of experimental conditions, that the phenomenon's robustness itself called out for analysis.

Eventually, almost 80 years after Brown first puzzled over the movements of his pollen granules, Einstein's 1905 work on molecular theory and thermodynamics provided the missing link, making sense of Brownian motion. The robustness of the Brownian motion pointed to the deeper, unobservable agitations of molecules constituting the medium – just as the evident rocking of a far off ship betrays imperceptibly distant waves on the sea (Perrin [1913], p. 83). As Perrin summarized, after reviewing the history of experiments testing the robustness of Brownian motion, "All of these characteristics force us to conclude [...] that the agitation does not originate either in the particles themselves or in any cause external to the liquid, but must be attributed to internal movements, characteristic of the fluid state" (Perrin [1913], p. 86).

This case exemplifies a general style of reasoning common across the sciences and much discussed in recent philosophy of science. To the extent that a scientific result is detected by numerous, diverse means, it is said to be *robust*. We can analyze the robustness of a particular result by exploring the differences in past means of detection that have had no (or a negligible) effect on the result – and also by exploring variations that *have* made a difference to the result. By performing such *robustness analysis* [henceforth, "RA"], we are often seemingly able to construct or confirm relevant scientific hypotheses. Regarding the above case, the Brownian motion is robust across various changes to the experimental apparatus (type of particle, medium, container, lighting) and sensitive to others (size of particle, temperature of medium). And it is upon analyzing such facts that Perrin says we are "forced to conclude," consistent with Einstein's molecular explanation, that there are internal, unobservable movements in the medium.

Robustness analyses are useful to science outside of the laboratory too; by "results" and "means of detection," I don't mean only to refer to experimental results and experiments. On the contrary, I mean these terms to be quite generic, so that 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. So understood, there are any number of example RAs from past and present scientific practice. These include the following cases discussed by philosophers of science:

Physics: In addition to the RA of Brownian motion already discussed, Perrin ([1913]) argues for the existence of atoms by noting that Avagadro's number is robust across thirteen distinct means of measurement (Cartwright [1991]; Salmon [1984]; Mayo [1996]). Hacking ([1983]) asserts that we lend further confirmation to the reality of microscopic phenomena when that phenomena is robustly observed through different types of microscope. See also Staley's ([2004]) discussion of RA's role in the discovery of the top quark.

Biological experiments: Culp ([1994]) discusses historical research on the bacterial mesosome, a "bag-shaped membranous structure" revealed in early electron micrographs. Scientists confirmed that mesosomes are artifactual when experiments showed that the presence of mesosomes depends on how a sample is prepared; i.e., mesosomes were not observed robustly across experiments using different sample preparation methods.

Psychology: Stolarz-Fantino et al. ([2003]) and Crupi et al. ([2008]) cite the "robustness," across diverse experimental setups, of the judgment that a conjunction is more likely than one of its conjuncts in order to argue for the reality of the conjunction fallacy.

Biological models: Weisberg and Reisman ([2008]) note that the 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") is robustly demonstrable across a variety of biological models, which differ in their simplifying assumptions. See also (Plutynski [2006]) for a discussion of RA in population genetics.

Climate science: Lloyd ([2010], p. 982) points to 14 different models used to simulate global climate. All of these models agree in that they include increases in greenhouse gas concentration as a "key causal factor" to global warming patterns. Lloyd argues that this allows us to "infer that greenhouse gas concentration increases cause global warming in the real world" – cf., (Parker [2011]). Pirtle et al. ([2010], p. 353) performed "a rough survey of the contents of six leading climate journals since 1990" and "found 118 articles in which the authors relied on the concept of agreement between models to inspire confidence in their results."

Economics Kuorikoski et al. ([2010]) and Odenbaugh and Alexandrova ([2011]) discuss work in which theoretical economists analyze the robustness of results derived from models differing in their idealizing assumptions. Kuorikoski et al. argue that the robustness of such results "guards us from error" by showing that the common results we get across such models "do not depend on particular falsehoods." Alternatively, Odenbaugh and Alexandrova argue that RAs have importance only in the context of discovery – not in contexts of theory justification.

One might roughly categorize the above into two groups, the first three being more empirically-driven and the latter three more analytically or model-driven. This would be consonant with a recent trend in philosophy of science. Breaking with the unifying focus on RAs *in general* that one finds, for example, in Wimsatt's work, many philosophers of science have found it useful to focus specifically on particular varieties of RA, allowing

them to examine potential idiosyncrasies of each type.[1] In this paper, however, I make a return to the general perspective and propose an account, which I claim unifies these types.

In more detail, the remainder of this paper provides a half critical, half constructive investigation into RA. The most important and challenging question an account of RA can answer is what it takes for various means of detection to be appropriately diverse. In Section 2, I argue that the prevailing formal explications of such diversity are unsatisfactory. Section 3 proposes a new unified account of RAs. At the heart of this account is an *explanatory* explication of what it takes for the relevant means of detection to be diverse. The result is, I suggest, truer to actual scientific RAs. Moreover, as I will argue with the help of some recent formal work on the "logic of explanatory power," this explanatory account affords us a helpful, new foothold on the logic undergirding RAs.

2 RA-Diversity and Independence

The most well-known early discussion of RA comes not from a philosopher of science, but from population biologist Richard Levins ([1966]). In a discussion of the various complications that confront biologists who use simplified models to study complex systems, Levins raises the daunting problem of how to decipher whether a result "depends on the essentials of a model or on the details of the simplifying assumptions." In this context, Levins ([1966], p. 423) famously proposes RA as a solution:

> [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.

Here, the common result is the similar behavior across various models – both Levins and Weisberg ([2006], [2013]) mention the similar qualitative behavior of predator-prey models interpreted as representing the Volterra Principle as an example. The diverse means which detect (i.e., entail) this result are the models themselves. Fitting with the informal characterization of RA that we have given, Levins notes that we may confirm that our theorem "depends on the essentials of a model" to the extent that it robustly follows from diverse models.

This brings us to the heart of the matter. The intuition underlying RA is that we can gain confirmation through diversity; certain hypotheses (e.g., the Volterra Principle) are supported to the extent that a result proves robust, and results are robust to the extent that we detect them in diverse ways. As Weisberg ([2013], p. 160) writes, "the key is to ensure that a sufficiently heterogeneous set of situations is covered in the set of models subjected to [RA]." But what precise sense of diversity is involved in RAs? What does it take for a set of situations to be "sufficiently heterogeneous"?

[1]For example, Woodward ([2006]) distinguishes four and Calcott ([2011]) three different types of RA. By contrast, Wimsatt ([2011], p. 299) clarifies that the aim of his earlier work on the topic was "to show the common features of these diverse methods."

Philosophers have offered various 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 diversity that drives our reasoning in RAs?

Any proposed account of such "RA-diversity" can be held accountable both to scientific practice and to our normative intuitions. Ideally, we would like a precise account that reveals the extent to which (and conditions under which) RA-diverse means of detection are able to lend confirmation to the relevant hypothesis. But such an account will 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 evidence.

In the remainder of this section, I apply this consideration in criticizing the most common formal approaches to explicating diversity in RAs. These Bayesian accounts model 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 many clear – even paradigmatic – cases of RA from science. To show this in each case, I return to the above examples of RAs in science – especially making use of the cases of Brownian motion and the Volterra Principle.

2.1 Unconditional probabilistic independence

So what precisely is meant, in RAs, when we say that the various means of detection are diverse? As a first attempt at answering this question, 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 simple (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 $Pr(R_j) = Pr(R_j|R_i)$.[2]

In their critique of Levins's discussion of RA, Orzack and Sober ([1993], pp. 539-40; cf., Justus [2012], pp. 798-99) 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."[3] In more detail and put more explicitly in Bayesian terms, when such a model implies a particular result in RA settings, we con-

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

[3]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."

sider it possible (and often even plausible) that the result is driven by the "essential core" of the model – i.e., Levins's common biological assumption(s). 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 (i.e., the fact that R is implied by one 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 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 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], p. 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.[4]

This *reliability independence* account differs from the simple account above. Instead of enforcing the overly stringent condition that the results of the various means of detection be probabilistically irrelevant to one another, this account just requires that if the means in question lead us astray in adopting some hypothesis, they do so for probabilistically independent reasons. Hence, 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], pp. 49-50) offers a quick probabilistic demonstration of this as follows: Assume that we have

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

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 giving us is incorrect, 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 clearly explicates an important notion of evidential diversity, it too will not do as a general explication of RA-diversity. Return to the example of Brownian motion. In this case, to say that an experiment is unreliable, or leading us astray, amounts to saying that it is misleading us in concluding with Perrin that there are internal, unobservable movements in the medium. Now consider again two of the RA-diverse means of detection (i.e., experiments) used by Brown: those in which a variety of pollen granules were suspended in water, and those in which a variety of inorganic materials were suspended in water. These experiments are cited by Perrin and others as diverse in the sense required for RA. Yet, credences about 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 – 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. But then learning that one of these experiments is leading us astray provides relevant information when deciding whether to trust the other. In particular, such information often should greatly reduce our confidence in the other's reliability.

This account faces the same problem with examples from modeling. In our example from biology, predator-prey models may be said to be unreliable to the extent that they mislead us in concluding that the Volterra Principle holds – e.g., if a model demonstrates behavior interpreted in accordance with this principle, but only because of some non-representative, artifactual assumption built into the model. Two RA-diverse predator-prey models that behave in accordance with 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., the unrealistic assumption that prey cannot take cover or learn). But then discovering that one of the models is unreliable should often 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 are similarly unreliable.

2.3 Confirmational and conditional independence

Lloyd ([2009], [2010]) has recently proposed another independence-based account. She proposes that RA-diversity amounts to *confirmational independence*, as explicated by Fitelson ([2001]). This sense is defined relative to a target 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 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 $c(H, R_j|R_i) = c(H, R_j)$.[5]

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], p. 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], p. 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."[6] By "screening-off," Fitelson has in mind the standard Reichenbachian ([1956], pp. 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*, 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

[5]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.

[6]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.

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 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 – i.e., if it's not true, all else being equal, that a general biocide will increase the abundance of the prey and decrease the abundance of predators when a two-species, predator-prey system is negatively coupled – there could be several reasons for why our models are displaying qualitative behavior interpreted in accordance with this principle. Perhaps this behavior is somehow 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. 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 (or reliability, or confirmational) independence. Wimsatt ([1981], p. 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."

I submit, however, that the criticisms above remain powerful even if we develop these accounts in this way. The problem 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. On the contrary, in certain such cases, 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 the foregoing examples, 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 way required 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.

3 Robustness Analysis as Explanatory Reasoning

In light of the last section, one might be tempted to conclude that there is no single, unified notion of RA-diversity. In RAs, we require our means of detection to be diverse, and this ambiguous requirement may be satisfied – one might say – to the extent that they are diverse in any of the above senses. Perhaps. But in the remainder of this paper, I try out a different sort of unified account of RA-diversity. It seems to me that philosophers working on RA have been lured away from the concept of RA-diversity by probabilistic independence; while this formal concept has been so fruitful in explicating notions of evidential diversity, I have argued that none of these notions of evidential diversity are the sort at work in RAs. The account developed here thus parts ways from independence-based accounts of evidential diversity, and instead has as its central notions *explanation* and *elimination*.[7]

3.1 Explanatorily discriminating means of detection

To motivate this account, it helps to work backwards, first thinking about what is accomplished in successful RAs by introducing diverse means of detection, and only then investigating the sense of diversity required to pull this off. It will again help to have in mind, as particular examples, Perrin's more experimentally-driven RA from Brownian motion to the hypothesis H that there are unobservable, "internal movements, characteristic of the fluid state" and a model-based RA supporting the Volterra Principle.

What work are the means of detection cited by Perrin supposed to accomplish by virtue of their diversity? What deficiencies would Perrin's RA have suffered if, for example, he had rested his case after reporting only that Brown observed various pollen granules in motion when suspended in water? The obvious answer is that, had Perrin stopped there, there would have too many competing potential explanations of the observed motion standing in the way of this conclusion. One would hardly be compelled to infer H from this single observation; Brown certainly was not. There were too many other hypotheses left standing at this stage for H (or any of H's competitors) to be well-supported by the result. So, instead of stopping there, Perrin mentions experiment after experiment detecting the result, which differ from one another insofar as they are able to eliminate different sets of H's competitors.

Similarly, a single mathematical model behaving in accordance with the Volterra Principle could hardly be said to lend strong support for this general biological principle. There are too many ways that any such model parts ways from the real world scenarios that it is trying to represent. The behavior of the model could be driven by any subset of the necessary idealizing and simplifying assumptions that we know fail to represent the real world. To rule out such competing explanations of the model's behavior, a variety of other models are thus employed, which likewise attempt to model behavior in

[7]The eliminative account of evidential diversity developed by Horwich ([1982], pp. 118-22) is a precursor to the present account. In (Schupbach [Forthcoming]), I discuss Horwich's account and its shortcomings as an explication of RA-diversity in detail.

accordance with the Volterra Principle but differ from one another on which simplifying and idealizing assumptions they involve.

As a first pass then, my proposal is that each newly cited detection in such cases *RA-differs* from previous means of detection if it is capable of ruling out another class of competing potential explanations of the result left standing by these previous means. In the case of the Volterra Principle, this amounts to showing model-by-model that the behavior of interest is not driven by various unrealistic assumptions. In the experimental case, in response to Brown's initial detection, H's critic asserts, "Well, perhaps this motion is really explained by the sexual drive inherent in pollen." So Perrin cites experiments showing that Brownian movements persist in other (organic) materials. "But perhaps the motion is due to a vital force in the material suspended." Perrin responds with experiments using inorganic materials. The imaginary dialectic proceeds until Perrin has cited experiments ruling out a host of competing explanations of the phenomenon. And it is only at the end of this process that Perrin says we are "forced to conclude" that H is correct.

The lessons we learn from these specific cases are generally applicable to all standard cases of RA – whether more empirically or analytically-driven. In any case, we can ask: What motivates us to seek and cite additional means of detection in an RA? What work are these additional means doing for us? The general account I am proposing responds that, at each increment of a RA, one cites an additional means of detection that has the power to discriminate between the target explanation of the result H and some competing potential explanation(s) H'. More specifically, such means of detection are *explanatorily discriminating between H and H′* in the sense that H would explain well our detecting result R via this new means, whereas H' would explain well our *failing to detect* R by this means.

Perrin's hypothesis that the Brownian motion is due to internal, invisible movements in the medium would explain well our observations of such motion using inorganic material, whereas the vital force hypothesis would explain well our failing to observe the result in this case; thus, this experimental detection explanatorily discriminates between these hypotheses. That a biological model behaves in accordance with the Volterra Principle while not making the unrealistic assumption that prey cannot take cover is explained well by the Volterra Principle itself (in conjunction with the hypothesis that the model is accurately modeling the real world behavior of predator-prey systems), but the competing explanation that this behavior is attributable to the unrealistic assumption in question would rather provide a strong explanation of our failing to observe the behavior using such a model; such a model thus explanatorily discriminates between these potential explanations.

The notion of explanatory discrimination provides us with a general (though still informal) way of characterizing RA-diversity:

RA-Diversity. Means of detecting R are *RA-diverse* with respect to potential explanation (target hypothesis) H and its competitors to the extent that their detections (R_1, R_2, ..., R_n) can be put into a sequence for which any member is explanatorily discriminating between H and some competing explanation(s) not yet ruled out by the prior members of that sequence.

In short, RA-diverse means of detection provide sequences of explanatorily discriminating bits of evidence, which successively eliminate more and more of H's competitors.

Importantly, on this account, it is really not so relevant whether means of detection are *strongly* diverse or sufficiently heterogeneous in some absolute sense, independent of considered hypotheses. What 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 the target hypothesis's salient competitors.[8]

As such, this account makes better sense of standard cases of RA in science. Many of the RA-diverse means of detecting Brownian motion cited by Perrin are really just not that different, and similarly for many of the predator-prey models used in the case of the Volterra Principle. Indeed, these means may be identical in all respects other than some modest change – e.g., in Perrin's case, in the particle suspended or mode of suspending it. This is why accounts that require RA-diverse means to be strongly different (often in a sense that pays no attention to the relevant hypotheses) run quickly into counterexamples in these cases.

Such means of detection are clearly explanatorily discriminating, 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 different than one another in some absolute sense, but because they are *relevantly* different than one another. The latter rules out a potential explanation 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 quite similar apart from some modest differences in their simplifying assumptions. But by utilizing these distinct (though perhaps overall quite similar) models, we may eliminate confounding explanations of our result left standing by either model used alone. As noted above, we may discard worries that our result is an artifact of a particular unrealistic assumption of the first model by using a second model that does not share that assumption.

Recall from Section 2 that there are two dimensions on which we might evaluate any proposed account of RA-diversity: fit with scientific practice and normative implications. I have just argued that the above explanatory account is truer to standard scientific cases of RA. But how does this account fare with regards to the second evaluative dimension? Is this account informative with regards to whether, under what conditions, and to what extent, RA is specially confirmatory? One might worry that the proposal, in virtue of its explanatory nature, is too vague to be of any normative service – that it trades off normative upshot for descriptive power. The remainder of this paper attempts to alleviate this worry. Recent work on the logic of explanatory power enables us to specify precise normative implications of RA-diverse (i.e., explanatorily discriminating) means of detection, and RAs generally.

3.2 The logic of robustness analysis

The intuition motivating any increment of a RA is that we can strengthen the case for our favored H by detecting our result using different means. Using our explanatory account

[8]This is not to say that the present account *requires* means of detection to be overall very similar to one another in some absolute sense; overall very different means of detection may of course be explanatorily discriminating with respect to H and its competitors.

of RA-diversity, we may clarify the following conditions for a successful increment of RA:

Past Detections. We have a result R that we have detected using $n-1$ different means. Let E denote the relevant conjunction $R_1\&R_2\&...\&R_{n-1}$.

Success. Hypothesis H (the hypothesis that we seek to confirm by our RA) explains this coincidence, but so does another hypothesis (or class of hypotheses) H'.

Competition. H and H' compete with one another, with respect to E.

Discrimination. There is an additional *explanatorily discriminating* (apropos H and H') means of potentially detecting R; i.e., in light of E, H would strongly explain our detecting R via this n'th means (R_n), whereas H' would strongly explain our not detecting R by this means ($\neg R_n$).

New Detection. The new means of detection concurs with prior means; i.e., R_n.

To explore the normative implications of RA, we first need to get more precise about some of the above conditions. First, consider **Competition**. What exactly does it take for two potential explanations H and H' to compete with one another?[9] The first thing to note is that there are deep, ontic senses in which hypotheses compete, but we are rather interested in the question of when hypotheses compete in our *epistemic* economy. In RAs, we consider hypotheses to be epistemic competitors in the sense that reason compels us to infer at most one of these; such epistemically competing hypotheses may or may not be competing in a deeper, ontic sense.

To see this important distinction, consider the most obvious sense in which hypotheses may compete: they may be mutually exclusive. While mutually exclusive hypotheses clearly compete ontically, they may or may not compete epistemically. The logical inconsistencies by virtue of which these hypotheses preclude one another may be tucked so subtly away into the fabric of the respective hypotheses, or the logical demonstration of such inconsistencies may be so computationally complex, that no rational person need recognize them. In other words, in principle, reason may not always oblige us to choose between jointly unsatisfiable hypotheses. Nonetheless, *if an agent does recognize that two hypotheses are inconsistent*, then that recognition will serve as reason compelling the agent to infer at most one of the hypotheses. So, where probabilities measure rational degrees of belief, we get the following epistemicized version of the mutual exclusivity sense of competition:

Competition (i). H and H' compete epistemically if $Pr(H\&H') = 0$.[10]

Is this sufficient condition for epistemic competition also necessary? In fact, no; it is easy to think of cases in which reason compels us to accept at most one of several potential explanations, despite the fact that they are recognizably consistent. There is no inconsistency in allowing that Brownian motion could simultaneously result from unobservable movements in fluid media *and* a vital force in organic matter. Nonetheless,

[9] Glass and Schupbach ([Unpublished]) offer a far more detailed account of hypothesis competition along these same lines—including a probabilistic explication of the degree to which hypotheses compete.

[10] We are assuming here and throughout that $Pr(H), Pr(H') > 0$, given that H and its competitors in the context of RAs are sufficiently plausible to worthy of consideration as potential explanations of R. Note that this assumption entails that when $Pr(H\&H') = 0$, these two hypotheses truly compete in the sense of precluding one another; the fact that their joint probability is zero is not due simply to the fact that H or H' takes probability zero on its own.

these hypotheses are viewed as epistemic competitors. An explication of epistemic competition that appeals to inconsistency will not then shed light on the sort of competition at work in all RAs.

Potential explanations of some explanandum E often compete, despite being consistent, when any one of these suffices to do the explanatory work of the others. Once we have accepted one, the explanatory work in accounting for E is done and hence there is no remaining explanatory reason from E to accept the others.[11] Perrin's hypothesis and its competitors compete in this way. If we favor Perrin's potential explanation of Brownian motion, then it would seem misguided to additionally accept the vital force hypothesis as another explanation. The phenomenon is already explained, so the explanandum no longer compels us to hunt for, and reason to, further explanations. In this case, the evidence compels us to accept at most one of the alternatives (which is precisely what we mean by epistemic competition). So we may clarify our second sense of epistemic competition as follows:

Competition (ii). H and H' epistemically compete in their potential explanations of E if H and H' both potentially explain E, but upon accepting H', H no longer retains any explanatory power over E.

Success, **Competition (ii)**, and **Discrimination** explicitly involve explanatory considerations. To make the logical implications of these conditions more precise, we can make use of recent work on the "logic of explanatory power". Schupbach and Sprenger ([2011]) recently develop and defend the following probabilistic measure of "the explanatory power that a particular explanans [h] has over explanandum [e]:"[12]

$$\mathcal{E}(e,h) = \frac{Pr(h|e) - Pr(h|\neg e)}{Pr(h|e) + Pr(h|\neg e)}$$

$\mathcal{E}(e,h)$ is a real-valued function with range $[-1,1]$. The greater the value of $\mathcal{E}(e,h)$, the stronger (more powerful) the potential explanation of e that h proffers. When $\mathcal{E}(e,h) = 1$,

[11] An anonymous reviewer helpfully observes that, depending on one's account of explanatory goodness, there may remain explanatory reason *apart from* E still supporting such hypotheses. For example, if a notion of simplicity – defined as a monadic property of hypotheses – counts toward explanatory goodness, then it may be that we no longer have explanatory reason from E for accepting a potential explanation, though we still have some explanatory reason apart from E for accepting this explanation (due to its great simplicity). The important point here is that, even if this is the case, the explanandum E itself will still only compel us to choose one of the hypotheses; insofar as we are inclined to go ahead and choose the other, it would be due for example to its simplicity and not to any support it retains from E.

[12] Probabilistic measures of explanatory power go back to Popper ([1959]) and Good ([1960]). More recently, in addition to Schupbach and Sprenger, McGrew ([2003]) and Crupi and Tentori ([2012]) propose candidate measures. All of these measures are explicitly meant to gauge the strength of the explanatory relation between an explanans and explanandum. While it is beyond the scope of the present paper to argue for measure $\mathcal{E}$'s special merits, the interested reader may refer to (Schupbach and Sprenger [2011]), which argues that this measure uniquely satisfies a set of intuitive conditions of adequacy related to the concept of explanatory power, and to (Schupbach [2011]), which demonstrates experimentally that this measure provides a close fit with our pre-formal intuitive judgments about explanatory power – at least in the tested, experimental contexts. Though I work directly with $\mathcal{E}$, all of the substantive results derived using this measure in this paper also hold using any of the alternative measures defended in the works cited above.

h provides a maximally powerful potential explanation of e, meaning that h suffices to account for (or predict) e with certainty. When $\mathcal{E}(e,h) = -1$, h provides a minimally powerful (maximally weak) potential explanation of e, meaning that h suffices to account for $\neg e$ with certainty. Finally, when $\mathcal{E}(e,h) = 0$, h is said to be explanatorily irrelevant to e, meaning that h accounts for e just as well as it accounts for $\neg e$.

To explicate **Success**, **Competition (ii)**, and **Discrimination** using $\mathcal{E}$, we formalize the judgment that h (potentially) explains e as a positive degree of explanatory power ($\mathcal{E}(e,h) > 0$), and we formalize the judgment that h strongly (potentially) explains e as near-maximal degree of explanatory power ($\mathcal{E}(e,h) \approx 1$). Statements about degree of explanatory power "in light of" or "upon accepting" some proposition p are naturally explicated using the following notion of *conditional* degree of explanatory power:

$$\mathcal{E}(e,h|p) = \frac{Pr(h|e\&p) - Pr(h|\neg e\&p)}{Pr(h|e\&p) + Pr(h|\neg e\&p)}$$

With the ideas of explanatory power and epistemic competition made more precise, we may now restate the conditions for a successful increment of RA formally:

Past Detections. We are given that $R_1\&R_2\&...\&R_{n-1}$; i.e., E
Success. $\mathcal{E}(E,H), \mathcal{E}(E,H') > 0$
Competition. **(i)** $Pr(H\&H') = 0$, or **(ii)** $\mathcal{E}(E,H|H') \leq 0$
Discrimination. $\mathcal{E}(R_n,H|E), \mathcal{E}(\neg R_n,H'|E) \approx 1$
New Detection. We learn that R_n

We are now in a position to investigate the normative import of a successful increment of RA. Below, I distinguish two cases for investigation, corresponding to the two senses in which hypotheses may epistemically compete.

3.2.1 Mutually exclusive competitors

Assume that H and H' compete in the sense of being recognized as mutually exclusive. In this case, we can partition the space of possibilities into the set $\{H, H', C\}$, where C is a catch-all hypothesis – logically equivalent to $\neg(H \vee H')$. To gauge the epistemic import of an increment of RA, we compare $Pr(H|E\&R_n)$ and $Pr(H|E)$ in order to clarify the conditions under which the former term might be greater than the latter, and what determines the extent of this inequality. We may use the above partition to expand $Pr(H|E\&R_n)$ and $Pr(H|E)$ using Bayes's Theorem and then compare the two by dividing the former by the latter:

$$\frac{Pr(H|E\&R_n)}{Pr(H|E)} = \frac{Pr(E\&R_n|H)}{Pr(E|H)} \times \frac{Pr(H)Pr(E|H) + Pr(H')Pr(E|H') + Pr(C)Pr(E|C)}{Pr(H)Pr(E\&R_n|H) + Pr(H')Pr(E\&R_n|H') + Pr(C)Pr(E\&R_n|C)}$$

Given that our n'th means of detecting R is explanatorily discriminating (as stipulated by **Discrimination**), we know that $\mathcal{E}(R_n,H|E) \approx 1$ and $\mathcal{E}(\neg R_n,H'|E) \approx 1$. When coupled with **Success**, these considerations entail $Pr(R_n|H\&E) \approx 1$ and $Pr(R_n|H'\&E) \approx 0$ respectively (proof in Appendix A).[13] From $Pr(R_n|H'\&E) \approx 0$, we know that $Pr(E\&R_n|H') = Pr(E|H')Pr(R_n|H'\&E) \approx 0$. These likelihoods

[13]Note that these likelihoods intuitively fit standard cases of RA. E.g., it is highly likely (indeed, nearly certain), that we should witness the Brownian motion using fragments of the Sphinx (under

accordingly reflect the intuition that explanatorily discriminating evidence that favors H over H' should effectively eliminate H':

$$\frac{Pr(H|E\&R_n)}{Pr(H|E)} = \frac{Pr(E\&R_n|H)}{Pr(E|H)} \times \frac{Pr(H)Pr(E|H) + Pr(H')Pr(E|H') + Pr(C)Pr(E|C)}{Pr(H)Pr(E\&R_n|H) + \cancel{Pr(H')Pr(E\&R_n|H')} + Pr(C)Pr(E\&R_n|C)}$$
$$\approx \frac{Pr(E\&R_n|H)}{Pr(E|H)} \times \frac{Pr(H)Pr(E|H) + Pr(H')Pr(E|H') + Pr(C)Pr(E|C)}{Pr(H)Pr(E\&R_n|H) + Pr(C)Pr(E\&R_n|C)}$$

$Pr(R_n|H\&E) \approx 1$ reflects the idea that H accounts so well for our robustly detecting R using this n'th means. Given this likelihood, we can show $Pr(E\&R_n|H) = Pr(E|H)Pr(R_n|H\&E) \approx Pr(E|H)$. This allows us to simplify the above equation further:

$$\frac{Pr(H|E\&R_n)}{Pr(H|E)} \approx \frac{Pr(H)Pr(E|H) + Pr(H')Pr(E|H') + Pr(C)Pr(E|C)}{Pr(H)Pr(E|H) + Pr(C)Pr(E\&R_n|C)} \quad (1)$$

Comparing like terms between the denominator and numerator of the right hand side of (1) (and remembering that the axioms of probability entail $Pr(E\&R_n|C) \leq Pr(E|C)$), we see that this ratio must be top heavy. Thus, in these circumstances, $Pr(H|E\&R_n) > Pr(H|E)$; a successful increment of RA will indeed confirm (increase the probability of) target hypothesis H to some extent.

What determines the extent? Examining the right hand side of (1) again, observe that the numerator of this ratio is greater than the denominator *at least* by a difference of $Pr(H')Pr(E|H')$. On the one hand, H''s prior probability $Pr(H')$ measures how plausible H's competitor was to begin with. On the other hand, $Pr(E|H')$ roughly measures competitor H''s fit with the evidence, prior to its being ruled out with the addition of R_n. These factors, considered together, provide an estimate of H''s overall epistemic virtue pre-elimination. So how good was H' (both on its own and in relation to E) before R_n? The answer to this question tells us how much better off (at least) H is now that we've ruled H' out. One might say that target hypothesis H soaks up the epistemic virtue that H' had going for it prior to being eliminated; spoils to the victor.

3.2.2 Consistent epistemic competitors

When H and its competitors are mutually exclusive, a successful increment of RA provides confirmation for H by ruling out certain ways in which H could be false. But what about cases in which H and its competitors are consistent with one another, but nonetheless compete in the sense of **Competition (ii)**? In such a case, by chopping away at H's competitors, explanatorily discriminating evidence also precludes possible ways in which H could be *true*. So it's not

background conditions specifying the proper setup of our experiment) given that the motion is due to unobserved agitations of the fluid medium, but this result is highly unlikely (indeed, it is nearly certain that it will not be observed) given that the motion is due to a vital force in organic matter.

obvious that RA will provide an effective strategy for confirming H in these cases.

Allowing for the possibility that H and H' are true together, we expand $Pr(H|E)$ in the following way:

$$\begin{aligned} Pr(H|E) &= Pr(H\&H'|E) + Pr(H\&\neg H'|E) \\ &= Pr(H|E\&H')Pr(H'|E) + Pr(H|E\&\neg H')Pr(\neg H'|E) \end{aligned} \tag{2}$$

(2) represents $Pr(H|E)$ as a weighted average of $Pr(H|E\&H')$ and $Pr(H|E\&\neg H')$. How do these compare to one another? Intuitively, in contexts of RA, the latter of these terms would seem to be the greater, reflecting the idea that E better supports H once competing hypotheses are ruled out.

Competition (ii) provides rational grounding for this intuition, revealing that $Pr(H|E\&H') <$ [or even $\ll$] $Pr(H|E\&\neg H')$. This condition implies $Pr(H|E\&H') \leq Pr(H|\neg E\&H')$; i.e., $Pr(H|E\&H')$ can be no greater than $Pr(H|\neg E\&H')$.[14] But in RAs, $Pr(H|\neg E\&H')$ will standardly be quite low indeed. This is the probability our target hypothesis takes in the case that all of our past evidence proves false and H's competitor proves true (e.g., the probability that unobservable agitations in fluid media cause Brownian motion conditional on us having repeatedly failed to ever observe such motion in past experiments, and conditional on such motion being attributable to a vital force in living matter). By contrast, $Pr(H|E\&\neg H')$ – representing the probability of H in light of all past evidence, and without H' standing in its way – can be considerable, and arguably high. Given these observations, $Pr(H|E\&\neg H')$ will typically provide a maximum cap (much greater than $Pr(H|E\&H')$) on the value of $Pr(H|E)$.

$Pr(H|E\&R_n)$ can also be expanded as follows:

$$\begin{aligned} Pr(H|E\&R_n) &= Pr(H\&H'|E\&R_n) + Pr(H\&\neg H'|E\&R_n) \\ &= Pr(H|E\&R_n\&H')Pr(H'|E\&R_n) + Pr(H|E\&R_n\&\neg H')Pr(\neg H'|E\&R_n) \end{aligned}$$

Recall from the previous section that **Discrimination**, when coupled with **Success**, entails that $Pr(R_n|H'\&E) \approx 0$. Consequently, it follows that $Pr(H'|E\&R_n) \approx 0$ – and so $Pr(\neg H'|E\&R_n) \approx 1$.[15] **Discrimination** thus again displays an eliminative effect in our analysis:

$$\begin{aligned} Pr(H|E\&R_n) &= \cancel{Pr(H|E\&R_n\&H')Pr(H'|E\&R_n)} + Pr(H|E\&R_n\&\neg H')Pr(\neg H'|E\&R_n) \\ &\approx \cancel{Pr(H|E\&R_n\&\neg H')} \end{aligned} \tag{3}$$

[14] This can easily be verified by stating the condition fully:

$$\mathcal{E}(E, H|H') = \frac{Pr(H|E\&H') - Pr(H|\neg E\&H')}{Pr(H|E\&H') + Pr(H|\neg E\&H')} \leq 0$$

[15] Given that $Pr(H'|E\&R_n) = Pr(H'\&E)Pr(R_n|H'\&E)/Pr(E\&R_n)$.

Does a successful increment of RA lend confirmation to H in these cases? We may specify the conditions under which it does, in light of the above, by comparing (2) and (3). Whether or not $Pr(H|E\&R_n) > Pr(H|E)$ comes down to whether

$$Pr(H|E\&R_n\&\neg H') > Pr(H|E\&H')Pr(H'|E) + Pr(H|E\&\neg H')Pr(\neg H'|E). \quad (4)$$

Recall that the right hand side of (4) is a weighted average of $Pr(H|E\&H')$ and $Pr(H|E\&\neg H')$, with a maximum cap value of $Pr(H|E\&\neg H')$. The strength of inequality (4) can thus be determined by asking two questions: 1. How much (if at all) greater than the right hand side's maximum cap $Pr(H|E\&\neg H')$ is the left hand side $Pr(H|E\&R_n\&\neg H')$? 2. How much (if at all) greater than the weighted average is this average's maximum cap – i.e., for fixed weights $Pr(H'|E)$ and $Pr(\neg H'|E)$, to what extent is $Pr(H|E\&\neg H')$ greater than $Pr(H|E\&H')$?

The degree to which R_n confirms H is thus determined by the strength of the following two inequalities: $Pr(H|E\&\neg H') > Pr(H|E\&H')$ and $Pr(H|E\&R_n\&\neg H') > Pr(H|E\&\neg H')$. This result can be intuitively interpreted and motivated. The first of these inequalities explicates one's judgment of how much more likely H is in light of E when H' is ruled out versus when H' is true. One might describe this as the evidential support that was previously not flowing from E to H due to the presence of competitor H'. We have already noted above that **Competition (ii)** gives us strong reason to think that this inequality will typically be quite significant indeed. The second inequality straightforwardly explicates the degree to which the new detection R_n confirms H, in light of the past evidence and H''s being eliminated. The degree to which RA supports H in this case then is determined by joining the support from past detections E that H' was previously cutting off from H to the additional support that R_n gives H in light of H''s elimination.

4 Conclusions

In this paper, I have argued against prevailing formal philosophical accounts of RA-diversity, and associated accounts of RA. Paradigmatic cases of RA, taken from the history of science, pose problems for views that require means of detection cited in RAs to be diverse in various senses to do with probabilistic independence.

In place of such accounts, I have developed a new explanatory account of RA-diversity and of RA. This account is descriptively superior to previous accounts, fitting nicely with actual cases of RA in the sciences. This account also has the virtue of being generally applicable across various subspecies of RA. In particular, I have suggested that the explanatory account of RA developed here applies to model-based RAs just as well as it does to empirically-driven RAs.

The explanatory account not only has these descriptive virtues, but when combined with recent formal epistemological work on the logic of explanatory power, it bears important normative fruit. There are at least two different senses in which explanatory hypotheses compete in RAs. Our investigation into the logic of RAs revealed that, in both cases, a successful increment of RA can indeed confirm a target hypothesis H.

Moreover, this investigation illuminated the conditions under which such confirmation occurs and the determinants of the degree to which the H is confirmed. In sum, we have found that, when RAs involve hypotheses that compete as mutually exclusive options, H is incrementally confirmed by ruling out possible ways that it could be false; and the extent to which it is confirmed is determined by how plausible the competing hypothesis was and how well that competitor fit the evidence pre-elimination.

When RAs instead involve consistent hypotheses that nonetheless provide competing potential explanations, H is confirmed by ruling out possible ways that H would lose its support from the evidence; and the extent to which it is confirmed is determined by adding in the support that it was not getting due to the presence of the competitor plus the support provided by the new, explanatorily discriminating means of detection. In either case, when RA is aptly described as a species of explanatory reasoning, its potential epistemic virtues are made manifest.

RA constitutes an intuitively compelling strategy, prevalent in the theoretical sciences, for constructing and confirming hypotheses. Philosophers of science today disagree, however, on whether it actually carries any normative, confirmatory power. The present work differs from recent accounts in its effort to articulate precisely the notion of diversity as it appears in actual scientific applications of RA. Recently, philosophers of science have either rested content with a vague (though still possibly somewhat informative) description of such diversity (e.g., Weisberg's description of such diversity as "sufficient heterogeneity"), or they have too readily leaned on Bayesian accounts of evidential diversity that ultimately do not correspond to the notion called upon in actual RAs. Here, I have developed a precise explanatory account of diversity, which I argue is descriptively truer to scientific practice. Only then have I investigated RA's acclaimed normative merits. Our precise account of the relevant concept of diversity reveals that RA can indeed be substantially confirmatory under certain natural conditions.

Appendix

Discrimination and **Success** entail $Pr(R_n|H'\&E) \approx 0$ and $Pr(R_n|H\&E) \approx 1$.

Discrimination requires that

$$\mathcal{E}(\neg R_n, H'|E) = \frac{Pr(H'|E\&\neg R_n) - Pr(H'|E\&R_n)}{Pr(H'|E\&\neg R_n) + Pr(H'|E\&R_n)} \approx 1.$$

Thus, $Pr(H'|E\&\neg R_n) - Pr(H'|E\&R_n) \approx Pr(H'|E\&\neg R_n) + Pr(H'|E\&R_n)$, which entails:

$$Pr(H'|E\&R_n) = \frac{Pr(H')Pr(E|H')Pr(R_n|H'\&E)}{Pr(E)Pr(R_n|E)} \approx 0. \tag{5}$$

Via **Success**, we are given that

$$\mathcal{E}(E, H') = \frac{Pr(H'|E) - Pr(H'|\neg E)}{Pr(H'|E) + Pr(H'|\neg E)} > 0.$$

This inequality is satisfied to the extent that

$$Pr(H'|E) = \frac{Pr(E|H')}{Pr(E)} > \frac{1 - Pr(E|H')}{1 - Pr(E)} = \frac{Pr(\neg E|H')}{Pr(\neg E)} = Pr(H'|\neg E);$$

or equivalently, to the extent that $Pr(E|H') > Pr(E)$. For this result to be consistent with (5) – so long as $Pr(H') \not\approx 0$, as we are assuming – it must be the case that $Pr(R_n|H'\&E) \approx 0$. □

Starting from $\mathcal{E}(R_n, H|E) \approx 1$ (which is also required by **Discrimination**), one may straightforwardly prove $Pr(R_n|H\&E) \approx 1$ *mutatis mutandis*.

Acknowledgements

I am grateful for the helpful conversations I have shared on this topic with Aki Lehtinen, Chiara Lisciandra, Gerhard Schurz, Jacob Stegenga, Ioannis Votsis, and fellow participants of a working seminar on robustness analysis organized by Chiara Lisciandra and hosted by the Finnish Centre for Excellence in the Philosophy of the Social Sciences (University of Helsinki, September 2014). Research for this article was supported by an Aldrich Fellowship from the University of Utah's Tanner Humanities Center, and was conducted during a visit to the Düsseldorf Center for Logic and Philosophy of Science (University of Düsseldorf).

Department of Philosophy
University of Utah
417 CTIHB, 215 S. Central Campus Drive
Salt Lake City, Utah, USA 84112
jonah.n.schupbach@utah.edu

References

Bovens, L. and Hartmann, S. [2003]: *Bayesian Epistemology*, New York: Oxford University Press.

Brown, R. [1827]: 'A Brief Account of Microscopical Observations on the Particles Contained in the Pollen of Plants; and on the General Existence of Active Molecules in Organic and Inorganic Bodies', Not Published.

Calcott, B. [2011]: 'Wimsatt and the Robustness Family: Review of Wimsatt's *Re-engineering Philosophy for Limited Beings*', *Biology & Philosophy*, 26(2), pp. 281–93.

Cartwright, N. [1991]: 'Replicability, Reproducibility, and Robustness: Comments on Harry Collins', *History of Political Economy*, 23(1), pp. 143–55.

Crupi, V., Fitelson, B., and Tentori, K. [2008]: 'Probability, Confirmation, and the Conjunction Fallacy', *Thinking and Reasoning*, 14(2), pp. 182–99.

Crupi, V. and Tentori, K. [2012]: 'A Second Look at the Logic of Explanatory Power (with Two Novel Representation Theorems)', *Philosophy of Science*, 79(3), pp. 365–85.

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, pp. 46–57.

Fitelson, B. [2001]: 'A Bayesian Account of Independent Evidence with Applications', *Philosophy of Science*, 68(3), pp. S123–40.

Glass, D. H. and Schupbach, J. N. [Unpublished]: 'When Do Hypotheses Compete?'

Good, I. J. [1960]: 'Weight of Evidence, Corroboration, Explanatory Power, Information and the Utility of Experiments', *Journal of the Royal Statistical Society. Series B (Methodological)*, 22(2), pp. 319–31.

Hacking, I. [1983]: *Representing and Intervening*, Cambridge: Cambridge University Press.

Horwich, P. [1982]: *Probability and Evidence*, Cambridge: Cambridge University Press.

Justus, J. [2012]: 'The Elusive Basis of Inferential Robustness', *Philosophy of Science*, 79(5), pp. 795–807.

Kuorikoski, J., Lehtinen, A., and Marchionni, C. [2010]: 'Economic Modelling as Robustness Analysis', *British Journal for the Philosophy of Science*, 61(3), pp. 541–67.

Levins, R. [1966]: 'The Strategy of Model Building in Population Biology', *American Scientist*, 54(4), pp. 421–31.

Lloyd, E. A. [2009]: 'Varieties of Support and Confirmation of Climate Models', *Proceedings of the Aristotelian Society, Supplementary Volumes*, 83, pp. 213–32.

Lloyd, E. A. [2010]: 'Confirmation and Robustness of Climate Models', *Philosophy of Science*, 77(5), pp. 971–84.

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

McGrew, T. [2003]: 'Confirmation, Heuristics, and Explanatory Reasoning', *British Journal for the Philosophy of Science*, 54, pp. 553–67.

Odenbaugh, J. and Alexandrova, A. [2011]: 'Buyer Beware: Robustness Analyses in Economics and Biology', *Biology & Philosophy*, 26(5), pp. 757–71.

Orzack, S. H. and 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), pp. 533–46.

Parker, W. S. [2011]: 'When Climate Models Agree: The Significance of Robust Model Predictions', *Philosophy of Science*, 78(4), pp. 579–600.

Perrin, J. [1913]: *Les Atomes*, Woodbridge, Conn: Ox Bow Press. Translated by D. Ll. Hammick.

Pirtle, Z., Meyer, R., and Hamilton, A. [2010]: 'What Does it Mean when Climate Models Agree? A Case for Assessing Independence Among General Circulation Models', *Environmental Science & Policy*, 13(5), pp. 351–61.

Plutynski, A. [2006]: 'Strategies of Model Building in Population Genetics', *Philosophy of Science*, 73(5), pp. 755–64.

Popper, K. R. [1959]: *The Logic of Scientific Discovery*, London: Hutchinson.

Reichenbach, H. [1956]: *The Direction of Time*, Berkeley, Cal: University of California.

Salmon, W. C. [1984]: *Scientific Explanation and the Causal Structure of the World*, Princeton: Princeton University Press.

Schupbach, J. N. [2011]: 'Comparing Probabilistic Measures of Explanatory Power', *Philosophy of Science*, 78(5), pp. 813–29.

Schupbach, J. N. [Forthcoming]: 'Robustness, Diversity of Evidence, and Probabilistic Independence', In U. Mäki, S. Ruphy, G. Schurz, and I. Votsis (*eds*), *Recent Developments in the Philosophy of Science: EPSA13 Helsinki*, Dordrecht: Springer.

Schupbach, J. N. and Sprenger, J. [2011]: 'The Logic of Explanatory Power', *Philosophy of Science*, 78(1), pp. 105–27.

Staley, K. W. [2004]: 'Robust Evidence and Secure Evidence Claims', *Philosophy of Science*, 71(4), pp. 467–88.

Stolarz-Fantino, S., Fantino, E., Zizzo, D. J., and Wen, J. [2003]: 'The Conjunction Effect: New Evidence for Robustness', *The American Journal of Psychology*, 116(1), pp. 15–34.

Weisberg, M. [2006]: 'Robustness Analysis', *Philosophy of Science*, 73(5), pp. 730–42.

Weisberg, M. [2013]: *Simulations and Similarity: Using Models to Understand the World*, New York: Oxford University Press.

Weisberg, M. and Reisman, K. [2008]: 'The Robust Volterra Principle', *Philosophy of Science*, 75(1), pp. 106–31.

Wimsatt, W. C. [1981]: 'Robustness, Reliability, and Overdetermination', In M. B. Brewer and B. E. Collins (*eds*), *Scientific Inquiry and the Social Sciences*, Jossey-Bass, pp. 125–63. 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), pp. 207–74. 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.

Wimsatt, W. C. [2011]: 'Robust Re-Engineering: A Philosophical Account?' *Biology & Philosophy*, 26(2), pp. 295–303.

Woodward, J. [2006]: 'Some Varieties of Robustness', *Journal of Economic Methodology*, 13(2), pp. 219–40.