"The Ravens Paradox" Is a Misnomer

Penultimate draft – please cite published version in Synthese.
The final publication is available at link.springer.com.

Abstract: I argue that the standard Bayesian solution to the ravens paradox—generally accepted as the most successful solution to the paradox—is insufficiently general. I give an instance of the paradox which is not solved by the standard Bayesian solution. I defend a new, more general solution, which is compatible with the Bayesian account of confirmation. As a solution to the paradox, I argue that the ravens hypothesis ought not to be held equivalent to its contrapositive; more interestingly, I argue that how we formally represent hypotheses ought to vary with the context of inquiry. This explains why the paradox is compelling, while dealing with standard objections to holding hypotheses inequivalent to their contrapositives.

Hempel's paradox of confirmation is about confirmation, not ravens. This sounds like a trivial observation, but it has been poorly appreciated in contemporary responses to Hempel's paradox. This is particularly true of the standard Bayesian solution, which crucially relies on the assumption that there are overwhelmingly more non-black things than ravens. The negative portion of this paper will be taken up with showing that the standard Bayesian solution is inadequate (though this should not be confused with attacking the Bayesian account of confirmation—for today, at least, I am a Bayesian myself), and that the literature has not produced a satisfactory response to the accusation of over-reliance on specific information about ravens. The positive portion of the paper will argue for another neglected idea: that a proper formalization of the hypothesis "All ravens are black," in the context of the paradox, would restrict its domain of applicability to ravens. The hypothesis would thus be inequivalent to its contrapositive, "All non-black things are non-ravens." That is not in itself a novel thesis; the really new idea in the positive part of this paper is that *how we should formally represent the hypothesis* *depends on the context of inquiry*. This is a matter of applied logic: formal systems are tools we use in representing hypothesis and evidence sentences; in different contexts, different tools will be appropriate, even for the same sentences. But before we get to any of that, we must have a statement of the paradox.

1. The Paradox

Here are four mutually incompatible statements, each of which is intuitively plausible.

> (1) (Equivalence Condition) If an evidence sentence E confirms a hypothesis H_1, then E also confirms any hypothesis H_2 equivalent to H_1 (and does so to the same degree).
>
> (2) (Nicod's Condition[1]) The hypothesis "All Fs are Gs" is confirmed by its instances, i.e., FGs.
>
> (3) The hypotheses "All ravens are black" and "All non-black things are non-ravens" are equivalent.
>
> (4) A non-black non-raven (e.g., a white shoe) does not confirm the hypothesis "All ravens are black."

In stating both (2) and (4), I have been wilfully sloppy, for the sake of convenience. Officially, confirmation is a relation among sentences, not between sentences and objects. For otherwise, as Hempel (1967) points out, we should have to deal with the problem of whether (assuming Nicod's Condition) a black raven with an albino sister confirms or disconfirms the hypothesis that all ravens are black. But it is sometimes easier to say, e.g., that a white shoe confirms the hypothesis than to say that the sentence "*a* is a white shoe" confirms the hypothesis, where "*a*" is a name of some object. I will buck official

[1] The name comes from an early proponent of the principle, Jean Nicod (1923).

practice when it is convenient.

(2) has had quite an adventure in the literature, and I shall not try to recount it all, though I shall cover some attempts to discredit the principle in §4. For now it is enough to point out that what one thinks of Nicod's Condition will depend on the contributions of background information. Many authors, Bayesians included, hold that confirmation is a three-place relation (between an evidence sentence, a hypothesis, and background information), and that Nicod's Condition does not hold against all possible backgrounds, so some restriction of the principle is in order. Most trivially, suppose our background information implies that if *a* is an FG, then something is an Fnon-G. Then if we find that *a* is in fact an FG, our background implies that the hypothesis "All Fs are Gs" is false, and so surely this instance of the hypothesis cannot be held to confirm it. Just what sort of background information would validate the principle is a thorny question, though. A reasonable interpretation of Hempel has it that Nicod's Condition is valid against tautological background information. This is naturally unsatisfactory to the Bayesian, who can make little sense of such an ultimately prior probability function. Maher (2004) uses a Carnapian framework to make sense of a probability function with tautological background information, but finds that Nicod's Condition is *not* valid in this case. (This contradicts his earlier (1999) analysis; I shall have more to say about Maher's supposed counterexample in §4.) But regardless of what one makes of such probability functions, it cannot be satisfactory to thus restrict Nicod's Condition: for we have more than trivial background information about ravens and their colouration, yet the paradox holds as described. There would be no paradox were Nicod's Condition not at least intuitively

attractive against our actual background information about ravens.[2] That is, if Nicod's Condition is construed as holding against any background information, it is stronger than we need to set up the paradox. All we need is an instance of the general version of Nicod's Condition: given what we ordinarily take for granted about ravens, finding a black raven confirms the hypothesis that all ravens are black.[3]

Now, I have said that the paradox is not really about ravens; what I mean is that (1)-(4) give an instance of a general phenomenon. We can give other instances of the same paradox without mentioning ravens or blackness. The challenge in understanding the paradox is to find what general characteristics of our information about ravens make the paradox compelling in this instance; those will likewise be the characteristics which make Nicod's Condition compelling. §2 will give an example of another instance of the paradox, this time not involving ravens, and §§3-4 will explore the status of Nicod's Condition in general, and the restrictions needed for it to be valid.

Finally, let us see why (1)-(4) are incompatible. Let *R* be the hypothesis "All ravens are black," *~B* the hypothesis "All non-black things are non-ravens," and *E* the sentence "*a* is a non-black non-raven," where "*a*" is a name of some object. By (2), *E*

[2] The point should not be overstated: some people may well find Nicod's Condition counterintuitive against any background information. E.g., certain Popperians who think a universally quantified hypothesis involves infinitely many objects will think the logical probability of the hypothesis will be zero given any finite evidence. The point here is that the intuition that Nicod's Condition is plausible against our ordinary background information about ravens is at least as widespread as the many authors who have taken the ravens paradox to be a genuine paradox—whatever proportion of these may have ultimately recommended rejecting Nicod's Condition. We cannot simply brush the paradox aside by rejecting Nicod's Condition out of hand.

[3] The truth of this assertion will depend on the referent of "we". I do not mean the people in question to have special information about, say, ravens' or birds' genetic makeup which might have as a consequence that the existence of a black raven makes the existence of an albino more likely; these, I take it, are not the inquirers normally imagined in setting up the paradox. Rather, I take the background information in question just to involve something like folk knowledge about ravens. Part of the task of §4 is to explore just what it is about the context of the paradox that makes it compelling; I take this to be relevant to the question of what background information is involved, but it is beyond the scope of this paper to determine precisely what that information is. Thanks to an anonymous referee for this point.

confirms ~*B*. By (3), *R* and ~*B* are equivalent, so by (1), *E* also confirms *R*. But by (4), *E* does not confirm *R*, which yields a contradiction.

2. The Standard Bayesian Solution

A satisfactory solution of the paradox will have to do several things:[4] (a) indicate which of (1)-(4) is false; (b) explain why the false statement is (or statements are) false; and (c) explain why the false statement is (or statements are) intuitively correct, despite its (or their) falsehood. For part (a), the standard Bayesian solution takes (4) to be false: a white shoe *does* confirm *R*; for part (b), there is a mathematical argument, which we shall come to presently. The rest of the solution comes in two flavours, one comparative and one quantitative. The comparative version has it that we mistakenly take (4) to be true because a white shoe confirms *R* less than a black raven; the quantitative version further claims that a white shoe confirms *R* only to a minute degree. (Both claims require endorsing some particular class of measures of degrees of confirmation.) Either way, the correct intuition that black ravens are more important for the investigation of *R* leads to the incorrect intuition that white shoes are irrelevant. Again, as with part (b), both versions of the solution to part (c) are supported by a mathematical argument, presented below.

For present purposes, we shall adopt the so-called difference measure of degree of confirmation. That is, the degree to which E confirms H relative to background information K is given by the difference P(H|E&K) – P(H|K).[5] Letting "R*a*", "B*a*", and *H*

[4] This set of requirements of a solution to the paradox is essentially what is found in Maher (1999: 51).
[5] With some minor additional provisos, the Bayesians' quantitative claim would still hold if we adopted instead any of the alternative measures described in Fitelson (1999), but the difference measure is the simplest to deal with.

respectively stand for "*a* is a raven," "*a* is black," and the hypothesis "(*x*)(R*x*→B*x*)",[6] we then have the following formal versions of the comparative and quantitative claims:

$$\text{(COMP) } P(H|{\sim}Ra\&{\sim}Ba\&K) - P(H|Ra\&Ba\&K) < 0$$

$$\text{(QUANT) } P(H|{\sim}Ra\&{\sim}Ba\&K) - P(H|K) \text{ is minute, but positive.}$$

Now, given only the assumption that the probabilities of ~R*a*&~B*a*, *H* and ~B*a*&*H* are all non-zero conditional on K alone, we have that:[7]

$$(*)\ P(H|{\sim}Ra\&{\sim}Ba\&K) - P(H|K) = P(H|K)\left(\frac{P({\sim}Ba|H\&K) / P({\sim}Ba|K)}{P({\sim}Ra|{\sim}Ba\&K)} - 1\right)$$

So far, all's well. But now the standard solution employs two questionable assumptions. First, we have Vranas's target in his (2004): that P(~*Ba*|*H*&K) is (approximately) equal to P(~*Ba*|K). From this, we can substitute 1 for the ratio P(~*Ba*|*H*&K) / P(~*Ba*|K) in (*). Second, we assume that, since there are overwhelmingly more non-black things than there are ravens, P(*Ra*|~*Ba*&K) is minute, say some very small ε. Then P(~*Ra*|~*Ba*&K) is 1 – ε, and so (*) reduces to:

$$P(H|{\sim}Ra\&{\sim}Ba\&K) - P(H|K) = P(H|K)\ [\varepsilon / 1 - \varepsilon].$$

The right-hand side of this equation will be positive so long as P(*H*|K) and ε are, and will be minute if ε is. (I.e., $\lim_{\varepsilon\to 0}(\varepsilon/1-\varepsilon) = 0$.) Thus, we have established (QUANT). To establish (COMP), it suffices to show (or assume, as we do here) that P(*H*|R*a*&B*a*&K) – P(*H*|K) is non-minute.[8]

But both of the assumptions I just called questionable are indeed questionable. As

[6] Readers with good memories may complain that I am multiplying symbols for this hypothesis beyond necessity, having already used *R*, but as will become clear in §3, I do not take the sentences I have represented with "*R*" ("All ravens are black") and with "*H*" to be equivalent in the context of the paradox.
[7] See Vranas (2004: 548, n9) for proof.
[8] Fitelson and Hawthorne (forthcoming: 13-15) provide an alternative argument for (COMP). That argument relies on the assumption that there are more (but not necessarily overwhelmingly more) non-black things than there are ravens. This assumption is as vulnerable to the arguments and counterexample to follow as its stronger counterpart in the argument for (QUANT).

Vranas (2004) argues, there is little support for the first assumption to be found in the literature;[9] but it is primarily with the second assumption that I make my quarrel here. The ravens paradox is not about ravens, and we cannot rely on facts about ravens which are contingent to the paradox. That is, we cannot rely on any facts about ravens which do not hold in other instances of the paradox. I claim that the assumption that there are overwhelmingly more[10] non-black things than ravens is such a contingent assumption.

Here, then, is an instance of the paradox where the assumption does not hold. As

---

[9] Though I think Vranas overstates the case against this assumption. The key passage in his attack is on p. 550 (emphasis in original): "Suppose one somehow refutes the claim that my estimate of the percentage of black objects should go up or down when I learn that they include every raven. The disputed assumption does not follow: it does not follow that my estimate should remain the same. What follows instead is that my estimate *may* remain the same." But I think a reasonable Bayesian could respond that, in the absence of positive reasons to do so, one ought not to change one's degree of belief in $\sim Ba$. Arguably, without some such principle of inertia, we might expect degrees of belief to vary wildly and unpredictably from moment to moment. Suppose a minute passes without my learning anything new. I have no positive reason to change my degree of belief in $\sim Ba$; but without a principle of inertia, we cannot say that my degree of belief should stay the same, but only that it may stay the same. Without a principle of inertia, we must admit that my degree of belief may instead go up or down (within the bounds set by my [unchanged] evidence). Such unprovoked changes in degree of belief may lead to erratic behaviour, if one's utilities stay fixed—and this seems an unpleasant consequence.

However, there is another reason to reject the independence assumption which Vranas does not mention. If $P(\sim Ba|H\&K) = P(\sim Ba|K)$, then we also have $P(H|\sim Ba\ \&K) = P(H|K)$. It then follows that $P(H|\sim Ba\&\sim Ra\&K) - P(H|\sim Ba\&K) = P(H|\sim Ba\&\sim Ra\&K) - P(H|K)$; that is, we get the same degree of confirmation from finding that an object selected completely at random is a non-black non-raven and from finding that an object selected at random *from the population of non-black objects* is a non-raven. (The argument holds for other measures of degree of confirmation. The result is immediate for the log-ratio and normalized difference measures; the log-likelihood-ratio measure requires a bit more work. We want to show that $\log[P(\sim Ra|H\&\sim Ba\&K) / P(\sim Ra|\sim H\&\sim Ba\&K)] = \log[P(\sim Ba\&\sim Ra|H\&K) / P(\sim Ba\&\sim Ra|\sim H\&K)]$. First, we have by Bayes's theorem that the argument of the log function on the left hand side is equal to $[P(H|\sim Ra\&\sim Ba\&K) \cdot P(\sim Ra\&\sim Ba|K) / P(H|K)] / [P(\sim H|\sim Ra\&\sim Ba\&K) \cdot P(\sim Ra\&\sim Ba|K) / P(\sim H|K)]$. This straightforwardly reduces to $[P(H|\sim Ra\&\sim Ba\&K) / P(H|K)] / [P(\sim H|\sim Ra\&\sim Ba\&K) / P(\sim H|K)]$. Next, take the argument of the log function on the right hand side of our goal equation, and apply Bayes's theorem to get $[P(H|\sim Ra\&\sim Ba\&K) \cdot P(\sim Ra|\sim Ba\&K) / P(H|\sim Ba\&K)] / [P(\sim H|\sim Ra\&\sim Ba\&K) \cdot P(\sim Ra|\sim Ba\&K) / P(\sim H|\sim Ba\&K)]$. By the independence assumption and some algebra, we get $[P(H|\sim Ra\&\sim Ba\&K) / P(H|K)] / [P(\sim H|\sim Ra\&\sim Ba\&K) / P(\sim H|K)]$, which is the same expression we got from the left hand side of the goal equation.)

I find this conclusion unacceptable, since I take the first object to be confirmationally irrelevant to the hypothesis, and the second to be confirmatory (given some further background information). The reason for the difference, in my view, is that there is a difference in the context of inquiry under which the hypothesis is to be considered, and thus a difference in how it is to be formalized. Contexts of inquiry are discussed more fully in §3.

[10] If, as the Popperians of n. 2 believe, there are infinitely many ravens and infinitely many non-black things under consideration here, we may not even be able to make sense of there being overwhelmingly more of one than the other. Since my point here is that it may not be true that there are overwhelmingly more non-black things than ravens, I am happy to concede this possibility. Thanks to an anonymous referee for this point.

in the standard ravens case, we consider a hypothesis of the form "All Fs are Gs." Where the standard case takes the Fs to be *ravens* and the Gs to be *black things*, let us now take the Fs to be *mammals*, and the Gs to be *things lacking a duckbill*. Our hypothesis, then, is "No mammal has a duckbill." The contrapositive—of the form "All non-Gs are non-Fs"—will then be "Everything with a duckbill is a non-mammal." The analogue of our previous assumption that there are overwhelmingly more non-black things than ravens (more non-Gs than Fs) would here be that there are overwhelmingly more creatures with duckbills than there are mammals, which is false. In fact, we can imagine a point in an imaginary evolutionary history where birds have evolved much later than mammals, so that there are overwhelmingly more mammals than ducks (and no platypuses, let's say);[11] in this case the standard Bayesian argument presented above would force us to conclude that a mammal with no duckbill only minutely confirms our hypothesis, but that a duck is quite significant. But the analogues of (3) and (4) here are both still intuitively compelling. (These would be, respectively, that our hypothesis and its contrapositive are equivalent, and that a non-mammal with a duckbill does not confirm our hypothesis. (1) and (2), Nicod's Condition and the Equivalence Condition, need no modification to apply to this case.) I see no intuitive reason to regard a duck as any more confirmationally relevant to a hypothesis about mammals than a white shoe is to a hypothesis about ravens. This is an instance of the ravens paradox without ravens, and one where the standard Bayesian solution helps not a bit.[12]

[11] Note that the important comparison here is not between the numbers of mammals and non-mammals, but between the numbers of mammals and duckbilled creatures.

[12] Note that if my target were Bayesianism, it would be promising to respond to this case simply by biting the bullet. That is, the Bayesian might just accept the counterintuitive conclusion that non-mammals are relevant to the hypothesis "No mammal has a duckbill," and so much the worse for our intuitions. But Bayesianism is *not* my target: I wish to claim not that Bayesianism is mistaken, but rather that the standard Bayesian solution to the ravens paradox does not, in fact, solve the paradox. The standard Bayesian

The observation that holding white shoes to be only minutely confirmatory of the ravens hypothesis relies essentially on a fact specific to ravens and blackness is not a new one—in fact one finds it as early as Hosiasson-Lindenbaum (1940: 140), the earliest presentation of the paradox in print!—but it has been only poorly dealt with. Hosiasson-Lindenbaum, for her part, points out that the assumption will generally hold for hypotheses of the form "All Fs are Gs," so long as F and G are both positive properties: for most such properties, there will be many more objects which lack the one property than possess the other. But even ignoring the difficulty of making a defensible metaphysical distinction between positive and negative properties, I see no reason for ruling out hypotheses about a mix of positive and negative properties. It seems to me that *being a mammal* and *lacking a duckbill* are prototypical positive and negative properties, respectively, but I should still like a theory of confirmation to encompass hypotheses like "No mammal has a duckbill."

We also have Fitelson and Hawthorne (forthcoming: 23-24), who attempt to avoid strong numerosity assumptions in their solution of the paradox. Instead of assuming that non-black things are much more numerous than ravens, they "[take] it as granted that *either* non-black things are much more numerous than ravens if [*H*] holds, *or* non-black things are much more numerous than ravens if [*H* does not hold]." But we are no better off with this assumption. Why should the truth or falsity of *H* have anything to do with how numerous ravens and non-black things are? Less nebulously, how do we know that ravens and non-black things are not approximately equinumerous, as may hold in the

solution hinges on explaining away our intuition that the white shoe is irrelevant to *R*; but the same argument will not explain our intuition that ducks are irrelevant to "No mammal has a duckbill." To bite the bullet here by accepting the counterintuitive conclusion in the mammal/duckbill case without explaining our intuitions against it is just to give up on solving the ravens paradox—though it is not to give up the case for Bayesianism. (Thanks to [redacted for blind review].)

mammal/duckbill case? I do not see that there is anything gained in the Fitelson-Hawthorne weakening of the standard assumption.

Hempel, who does not accept the standard Bayesian solution, does note that the assumption about how many more non-black things than ravens there are will not hold, in general. "[I]t remains an empirical question, for every hypothesis of the form 'All *P*'s are *Q*'s", whether actually the class of non-*Q*'s is much more numerous than the class of *P*'s; and in many cases this question will be very difficult to decide" (1945: 21-22, n2). He has little more to say on the subject than this—which is not terribly surprising, since the paradox had not yet generated the mountainous literature that was to come—and it is unfortunate that his observation went largely unheeded in the years to follow.

3. Inequivalence of the Contrapositive, and Flexibility of Representation

Parts (a) and (c) of the solution I favour are easily stated. I claim that (3) is false: *R* ("All ravens are black") is not equivalent to its contrapositive *~B* ("All non-black things are non-ravens") in the context of the paradox, for reasons I shall come to presently. The intuitive appeal of (3) is easy to explain for anyone who has taken an introductory course in logic. We standardly formalize universal generalizations like *R* with sentences of the form "$(x)(\mathrm{R}x \rightarrow \mathrm{B}x)$", which is classically equivalent to the parallel formalization of *~B*, "$(x)(\sim\mathrm{B}x \rightarrow \sim\mathrm{R}x)$". With all the valuable advances we have made thanks to classical logic, coupled with the standard formalizations of ordinary language taught in introductory logic classes, it is no surprise that we should be inclined to apply those same methods to the hypothesis *R*.

Despite the strong *prima facie* case for representing *R* in the standard way, I think this move is the misstep that leads to paradox. The reason it is unintuitive that a white shoe should confirm *R* is that *R* seems to be *about* ravens. Shoes are irrelevant. (Likewise, ducks have nothing to do with a hypothesis about mammals.) The standard formalization of *R* is unfaithful to this intuition. Rather than saying something only about ravens, it says something about all objects—viz., that they are not non-black ravens. This opens the door to paradox, and we might as well close the problem here. I propose we represent *R* instead as "(x)(Bx)," or in other words, "For all x, x is black," where the quantifier is understood to range over the class of ravens.

I am hardly the first to adopt the general strategy of denying the equivalence of *R* and ~*B*, but suggestions of alternative representations for *R* are usually more radical than mine. For example, Sylvan and Nola (1991) replace the material conditional with a relevant conditional; and Cohen (1987) uses a three-valued logic, defining a presupposition operator "*", such that "*Ra→Ba" is neither true nor false if a is not a raven. But I think my suggestion has two significant advantages over these alternatives. First, one need not give up classical logic to use my representation of *R*. There are those who would say this is nothing to crow about—classical logic is an unpopular candidate for the One True Logic in certain circles—but it is undeniably easy to work with. I do not claim to be giving the underlying logical structure of all universal generalizations, nor to have deep insights into the use of conditionals in ordinary language. I am only giving what I see as a useful formal representation of a certain type of hypothesis *in a certain context of inquiry*, which yet remains faithful to our intuitions. There is a definite pragmatic advantage to a solution to the paradox which does not require radical logical reform. Besides, while classical logic arguably has its paradoxes, relevance and three-valued logics have their own unintuitive quirks: for example, disjunctive syllogism is relevantly invalid.

This brings us to the second, and more significant, advantage of my proposal. The idea of solving the ravens paradox by taking *R* to be restricted to a certain domain of application is not at all new. Hempel (1945: 17-18) considers this possibility in his presentation of the paradox, and he gives an important objection to it, to which Sylvan and Nola and Cohen are vulnerable, but my proposal is not. Hempel gives the following example. Suppose we have a piece of some salt whose composition is unknown. Suppose also that the hypothesis that all sodium salts burn yellow is very well confirmed. We heat our piece of salt and find that it burns blue; we conclude that it is not a sodium salt. It looks like in this case we have used the contrapositive of our hypothesis. In light of the hypothesis, we move directly from “This does not burn yellow” to “This is not a sodium salt.” This inference is not permissible from our hypothesis alone if we take it to be inequivalent to its contrapositive. The lesson here is that *sometimes* we do use universal generalizations as if they were equivalent to their contrapositives. Accounts such as Cohen’s and Sylvan and Nola’s, which treat all such generalizations alike, irrespective of the context of inquiry, cannot be satisfactory. Cohen (1987: 159) even goes so far as to say that the form of a hypothesis such as “All sodium salts burn yellow” incorporates a testing procedure; but Hempel clearly shows that that same hypothesis can be used in all sorts of wildly different experiments.

My account does not require such rigidity of representation. In some contexts, *R* ought to be taken as relevant only to ravens; in others, it will say something about a wider class of objects. To take another example, Sainsbury (1995: 81-82) gives us the following. There has been an outbreak of Legionnaires’ disease, and

> Our hypothesis is that the source of the infection was the water at St George's school, consumed by all the children who attended last week. Will only an instance of the generalization "All pupils attending St George's last week contracted Legionnaires' disease" confirm it? Imagine that we find some St George's children who are free from the disease, but that it then turns out they missed school last week. We would normally count this as evidence in favour of our hypothesis … and yet these children are not instances of the hypothesis.

First, I would note that despite the fact that the hypothesis in Sainsbury's example ought to be held equivalent to its contrapositive, this is so *only within a certain relevant domain*. For example, supposing St George's school is in Vancouver, we should consider it confirmationally irrelevant if someone in, say, Mongolia is found not to have contracted Legionnaires' disease, and not to have attended St George's last week. The hypothesis ought to be taken to apply to those whom we might expect to have attended St George's last week: the pupils, teachers, and staff. But I do agree that the hypothesis here ought to be held equivalent to its contrapositive within that domain. This is because we are dealing with a context of inquiry quite different from that of the ravens paradox. Here we are investigating the cause of the outbreak of Legionnaires' disease. We are not merely enumerating the properties of the pupils who attended St George's last week; we are not interested in the class of attending pupils for its own sake, but because we suspect that learning about them will tell us about the spread of the disease. The reason we are interested in pupils who did not attend St George's last week is that information about them will tell us more about the outbreak. Our Mongolian tells us nothing about the outbreak, and so is counted irrelevant.

Hempel's and Sainsbury's examples show that there are contexts where we must either hold universal generalizations are equivalent to their contrapositives or do violence to our intuitions. The ravens paradox is another story. There would be no paradox if it were not counterintuitive to suppose that white shoes are confirmationally relevant to a hypothesis about ravens. Therefore, I claim that an acceptable solution to the ravens paradox must hold *R* inequivalent to its contrapositive in the context of the paradox, but that we must allow the equivalence to hold in other contexts. It remains, then, to say something about what sort of contexts will yield the equivalence of *R* and ~*B*. We have now three examples, in two of which a hypothesis of the form "All Fs are Gs" is to be equivalent to its contrapositive. These are, first, Hempel's case where the hypothesis is employed to determine whether a given object is an F, having found that it is a non-G; and, second, Sainsbury's case where we are investigating a causal connection between Fs and Gs. We have seen that these contexts give us reason to take the hypothesis to be saying something about non-Fs as well as Fs. The context of the ravens paradox can, I think, be characterized as *investigating the properties of the Fs*. If we are only interested in the properties of the Fs, we have no reason to take the hypothesis to be saying anything about non-Fs.

But this answer only pushes the question back one step. We want to know what it is about the ravens paradox that makes it desirable to hold *R* inequivalent to ~*B*; the answer that the paradox holds in a context where we are investigating the properties of ravens leaves us to ask why this should be so. For there is nothing explicit in the statement of the paradox itself to indicate any such practice. All we are given is a hypothesis, and somehow this is enough for us to rule it unintuitive that a certain procedure (turning up non-black non-ravens through random sampling) should produce confirming instances of the hypothesis. If this intuition is to be explained through appeal to some context of inquiry, there must be something in the statement of the hypothesis to make us assume that we are dealing with the context in question. For there are certainly other contexts than investigating the properties of ravens where the question of *R*'s truth might arise. For example, analogously with Hempel's sodium salts example, if we know that all ravens are black, and we are tracking some bird which might be a raven, then from finding that the bird has dropped a pink feather, we can infer that the bird is not a raven. This is a context where we should hold *R* equivalent to ~*B*.

I think what makes us assume that *R* is to be understood in the context of investigating the properties of ravens is that ravens are a natural kind, in Quine's (1969) sense. (Or at least they seemed so to Quine; the following remarks still hold if we take species to be individuals rather than natural kinds, as many philosophers of biology now do.) Quine appealed to the notion of natural kinds to explain away both the ravens paradox and Goodman's grue paradox, by claiming that only natural kinds are "projectible," i.e., "All Fs are Gs" is confirmed by an FG only if Fs are a natural kind. This strategy has been rightly criticized by Chihara (1981), but I think Quine was not far wrong. I want to make a weaker claim about natural kinds and confirmation: I say only that, with hypotheses of the form "All Fs are Gs," if Fs are a natural kind, then there is a default assumption that we should think about a ravens-paradox-like context of inquiry when asking ourselves whether various instances (FGs, non-F non-Gs, etc.) are confirmatory, disconfirmatory, or irrelevant. Natural kinds carve up the world in ways we find natural; they are the sorts of things we might be interested in investigating just to explore what they are like. (Likewise, if there is an individual we find particularly interesting or important, we may well be interested in a catalogue of its properties, just to find out what it is like.) It is not strange for a zoologist, say, to study the members of some understudied species without intending to test any particular hypothesis, but simply to discover how they behave. (Likewise, an astronomer studying a newly-discovered planet.) Perhaps closer to the official story of the ravens paradox—which involves random sampling from a population of ravens and non-ravens, unlike the case of the zoologist who actively selects which organisms to observe—before we ever had scientists, we had a wealth of observations about our natural surroundings, gathered just by living among ravens and rocks and rainstorms. Wondering whether all ravens are black, it is not unreasonable to appeal to common experience for confirming instances: "Ravens are common, and we have never seen a non-black one," one's fellows might say. Here the confirming observations (black ravens) were not selected for their ravenhood or their blackness. Rather, they are basically random observations, coming in alongside observations of rocks and trees and pigeons. But because ravens are a natural kind, we can organize this mass of basically random observations to come up with useful information about the restriction of these observations to ravens. With non-natural kinds like non-black things, I doubt the same thing would be possible. Likewise, scientists are unlikely to actively follow a group of non-black things in order to discover properties of non-black things in general. Quine may have been wrong to suppose that there is a logical problem with taking instances of non-black non-ravens to confirm ~*B*, but I suggest there is a (conventional) default assumption that "All Fs are Gs" is to be taken in a context of exploring the properties of the Fs if the Fs are a natural kind. I do not claim there is anything logically or empirically special about natural kinds, just that there is something psychologically special about them: we understand hypotheses about natural kinds differently from hypotheses about non-natural kinds.

4. Nicod's Condition Revisited

The solution I have been defending involves holding that Nicod's Condition is valid, at least in the context of the ravens paradox. At the end of the previous section, I described conditions I think are sufficient for the paradox to hold, and which therefore ought to be sufficient for Nicod's Condition to hold. But the conditions I gave were negative rather than positive: if we have a hypothesis of a certain form, describing a natural kind, then there is a default assumption that we are investigating the properties of the objects the hypothesis is about. That is, *in the absence of any further background information*, we should assume we are in such an exploratory context. So Nicod's Condition should hold for hypotheses of the appropriate form, so long as they are about natural kinds, given no further background evidence. It is time, then, to consider some of the supposed counterexamples to Nicod's Condition found in the literature. I find that these fall into two basic groups: (a) those relying on illicit information in the evidence sentence, and (b) those relying on illicit background information.

Type (a) examples could also be explained as mistaking confirmation as a relation between hypotheses and objects, or states of affairs, rather than between hypotheses and evidence sentences. We have already seen the difficulties of such a position in §1. For example, here is a supposed counterexample to Nicod's Condition of type (a), based on Swinburne (1971):

> Let *h* assert that killer bees cannot live north of the 38th parallel north. Let *e* assert that killer bees have been observed at 37.99°N. This is a positive instance of the hypothesis that *disconfirms h*. (Korb 1993: 146)

The trouble here is that *e* says too much. Remember that Nicod's Condition says that evidence of the form "*a* is an FG" confirms a hypothesis of the form "All Fs are Gs."

With *h* as given in the example, we must take the Fs to be killer bees, and the Gs to be things that cannot live north of 38˚N. Then Nicod's Condition says only that the evidence sentence *e´*, "*a* is a killer bee who cannot live north of 38˚N," confirms *h*; but *e* asserts more, that "*a* is a killer bee who lives at 37.99˚N." If we have this further information about where exactly *a* does live, it becomes more likely that there will be another killer *b* who lives very slightly north of *a*, across the 38th parallel north. But without that added information, it is perfectly plausible to hold that *h* is confirmed by *e´*, and this is all Nicod's Condition requires.

Now, if we held that objects or states of affairs confirmed, rather than sentences, it would be understandable why we should not make the distinction between *e* and *e´*. For considering the bee itself who lives at 37.99˚N, or the state of affairs of having a bee live there, would be enough to make us lower our probability of *h*. But this is an untenable view of confirmation, as we saw in §1.

Another example of type (a) comes from Rosenkrantz (1977: 35). Three philosophers leave a party, picking up their hats from the closet on their way out. The closet was dark, and so each picked a hat at random. (They were the only three with hats at the party.) Philosopher 1 forms the hypothesis, "Each philosopher has someone else's hat." As evidence, she sees that Philosopher 2 has Philosopher 3's hat, and Philosopher 3 has Philosopher 2's hat. These are both instances of the hypothesis, and so she takes them to confirm her hypothesis. But clearly, given these two pieces of evidence, the hypothesis must be false.

The trouble, again, is that the evidence sentences are too precise. Nicod's Condition only tells us about the confirmational import of sentences of the form

"Philosopher *n* has someone else's hat." It's just greedy to ask exactly whose hat Philosopher *n* has. Again, if we thought that the hats did the confirming, rather than sentences about the hats, we would be tempted to reject Nicod's Condition here.

As an exemplar of type (b), we have Maher's (2004: 77-78) numerical example. Maher takes a Carnapian approach to confirmation, giving a series of axioms a satisfactory probability function must obey, going beyond the axioms of probability calculus. (This is a departure from the subjective Bayesian approach, which holds that the only rational constraints on a subjective probability function are the probability calculus and the rule of updating by conditionalization. Most of Maher's extra axioms are to be found elsewhere, in his (2000).) There is also the further, rather vague, injunction that determining the prior probability that *a* is a raven, for example, "would require careful consideration of what exactly is meant by 'raven' and what the alternatives are" (2004: 75). Thus, despite Maher's desire to get his probabilistic results from a system of inductive logic, the determination of prior probabilities—which will be important in his supposed counterexample to Nicod's Condition—seems to require some empirical knowledge about the properties concerned.

I will not run through Maher's example in numerical precision, for that would require proving some theorems of his system, which is outside the scope of this paper. But we need not be so precise here. Our hypothesis, as usual, is "All Fs are Gs." We have only tautological background information. However, we know that Fs and Gs are both quite uncommon, and Gs are far less common than Fs. Then, since it is unlikely there are any Fs at all, the prior probability of our hypothesis will be high. For if there are no Fs, then "All Fs are Gs" is true. But if we find, at random, an FG, then this will raise the probability that there are other Fs (because of analogy effects built into Maher's probability functions), as well as the probability that there are other Gs. But since Gs are much rarer than Fs, it is not unlikely that, if there is another F, it is a non-G. Thus the probability of our hypothesis will drop; it is disconfirmed by finding an FG.[13]

I promised a vague presentation of Maher's case, and I think I have delivered just that; but the numbers do bear out essentially the effects I have described. Still, there is enough here to complain about. Maher's example requires setting the prior probabilities of an object being an F and being a G very low; but what could justify such an assignment except some information about the Fs and the Gs? Surely this would require some non-tautological background information; and this is exactly what Maher needed to avoid using to make his point. To be sure, the "background information" space in his formalism is empty, but it is only in a purely technical sense that he has shown Nicod's Condition to be invalid against tautological background. Rather, he has shown that Nicod's Condition fails against the background information encoded in the prior probabilities given.

Conclusion

I have argued that the standard Bayesian solution to the ravens paradox is insufficiently general to be satisfactory. I gave an example of an instance of the paradox to which the standard Bayesian solution gives the wrong answer. I argued that we should instead formalize hypotheses differently in different contexts of inquiry, and in particular that we should formalize *R*, in the context of the paradox, as "(*x*)(B*x*)," with the quantifier

[13] Maher (2004: 78, n4) notes that this example is quite similar to Good's (1968) reply to Hempel (1967). The latter paper pointed out that Good's (1967) counterexample to Nicod's Condition relied on illicit background information, whereas Hempel only held that the Condition is valid against tautological background.

understood to be restricted to ravens. But the main positive point I want to make is not that this is the right formalization of *R* in the context of the paradox—I am friendly to efforts along the lines of Cohen (1987) or Sylvan and Nola (1991). Rather, my main claim is that there is no single correct formal representation of a given hypothesis that will be appropriate in all contexts. The wrong representation can lead to paradoxes like that of the ravens.

It bears emphasizing that I do not have anything to say about the logical structure of hypotheses like "All ravens are black." I am especially not defending any sort of contextualist semantics. Although I have been arguing that when we use the sentence "All ravens are black" in different contexts of inquiry, we may need different formal representations of that sentence to best bring out the relations of confirmation in each context, it is important that the hypothesis sentence mean the same thing across contexts. After all, the reason we can legitimately assume, for example, that all sodium salts burn yellow in Hempel's imaginary experiment is because we have confirmed that hypothesis in other contexts of inquiry, where we may have been using a different formal representation of the (same!) hypothesis in determining what confirms what to what degree. I have nothing interesting to say here about meanings.

Finally, I want to point out that although I have attacked the standard Bayesian solution to the paradox, and I have not used Bayesian techniques in my own solution, I am not attacking Bayesianism; I think the Bayesian explication of confirmation is as close to correct as any of its rivals. There is no reason why Bayesians should be unable to use different formal representations of the same hypothesis in different contexts of inquiry. My proposed solution to the paradox gives no reason to choose between competing general theories of confirmation.[14]

References

Chihara, Charles (1981). "Quine and the Confirmational Paradoxes," in P.A. French, T.E. Uehling Jr. and H.K. Wettstein (eds.), *The Foundations of Analytic Philosophy* (*Midwest Studies in Philosophy*, vol. 6). Minneapolis, MN: University of Minnesota Press, pp. 425-52.

Cohen, Yael (1987). "Ravens and Relevance," *Erkenntnis* 26: 153-79.

Fitelson, Branden (1999). "The Plurality of Bayesian Measures of Confirmation and the Problem of Measure Sensitivity," *Philosophy of Science* 66 (Proceedings): S362-78.

Fitelson, Branden and James Hawthorne (forthcoming). "How Bayesian Confirmation Theory Handles the Paradox of the Ravens," in E. Eells and J. Fetzer (eds.), Probability in Science, Open Court. Page references are to .pdf file from <http://fitelson.org/research.htm>.

Good, I.J. (1967). "The White Shoe is a Red Herring," *British Journal for the Philosophy of Science* 17: 322.

Good, I.J. (1968). "The White Shoe *Qua* Herring Is Pink," *British Journal for the Philosophy of Science* 19:156-7.

Hempel, Carl G. (1945). "Studies in the Logic of Confirmation," *Mind* 54: 1-26, 97-121

Hempel, Carl G. (1967). "The White Shoe: No Red Herring," *British Journal for the Philosophy of Science* 18: 239-40.

Hosiasson-Lindenbaum, Janina (1940). "On Confirmation," *Journal of Symbolic Logic* 5: 133-48.

Korb, Kevin B. (1993). "Infinitely Many Resolutions of Hempel's Paradox," in R. Fagin (ed.), *Theoretical Aspects of Reasoning about Knowledge: Proceedings of the Fifth Conference*. Asilomar, CA: Morgan Kaufmann, pp. 138-49.

Maher, Patrick (1999). "Inductive Logic and the Ravens Paradox," *Philosophy of Science* 66: 50-70

Maher, Patrick (2000). "Probabilities for Two Properties," *Erkenntnis* 52: 63-91.

Maher, Patrick (2004). "Probability Captures the Logic of Scientific Confirmation," in C. Hitchcock (ed.), *Contemporary Debates in the Philosophy of Science*. Oxford: Blackwell, pp. 69-93.

Nicod, Jean (1923). *Le Problème Logique de l'Induction*, Paris: Alcan.

Quine, W.V.O. (1969). "Natural Kinds," in *Ontological Relativity and Other Essays*. New York: Columbia University Press, pp. 114-138.

Rosenkrantz, Roger (1977). *Inference, Method and Decision*. Boston, MA: Reidel.

Sainsbury, R.M. (1995). *Paradoxes*, second edn. Cambridge: Cambridge UP.

Swinburne, R.G. (1971). "The Paradoxes of Confirmation—A Survey," *American Philosophical Quarterly* 8: 318-30.

Sylvan, Richard and Robert Nola (1991). "Confirmation without Paradoxes," in G.

[14] For comments on earlier versions of this paper, I am grateful to Murat Aydede, Peter Brössel, Margaret Schabas, Roger Stanev, and especially to Paul Bartha and Chris Stephens. Parts of this paper were presented to the 2009 Pacific APA, the 2009 Society for Exact Philosophy, and the UBC Graduate Colloquium; I am grateful to the audiences for helpful discussion.

Schurz and G.J.W. Dorn (eds.), *Advances in Scientific Philosophy: Essays in Honour of Paul Weingartner*. Amsterdam: Rodopi, pp. 5-44.

Vranas, Peter B.M. (2004). “Hempel’s Raven Paradox: A Lacuna in the Standard Bayesian Solution,” *British Journal for the Philosophy of Science*55: 545-60.