

Forthcoming in *Philosophy of Science*. Penultimate version.

A New Bayesian Solution to the Paradox of the Ravens¹

Susanna Rinard

Abstract

The canonical Bayesian solution to the ravens paradox faces a problem: it entails that black non-ravens disconfirm the hypothesis that all ravens are black. I provide a new solution that avoids this problem. On my solution, black ravens confirm that all ravens are black, while non-black non-ravens and black non-ravens are neutral. My approach is grounded in certain relations of epistemic dependence, which, in turn, are grounded in the fact that the kind *raven* is more natural than the kind *black*. The solution applies to any generalization “All *F*’s are *G*” in which *F* is more natural than *G*.

1. Introduction

Black ravens provide some evidence that all ravens are black, but non-black non-ravens do not. Why is this, given that all ravens are black if and only if all non-black things are non-ravens? I will present a new Bayesian solution to this much-discussed paradox.

¹ Many thanks to Mark Crimmins, Adam Elga, Peter Godfrey-Smith, Andrew Graham, Edward Hall, Alan Hajek, Agustin Rayo, Robert Stalnaker, Michael Strevens, Roger White, Stephen Yablo, three anonymous reviewers for *Philosophy of Science*, and an audience at the 2007 MITing of the Minds conference. Special thanks to Miriam Schoenfield for detailed comments on multiple drafts. I am also very grateful to Vann McGee for answering some technical questions.

A new approach is needed because the canonical Bayesian solution faces a serious problem. It relies on two independence assumptions that jointly entail that black non-ravens disconfirm the hypothesis that all ravens are black. This is problematic because intuitively, black non-ravens are just as irrelevant as non-black non-ravens.

I begin the case for my alternative approach by showing that if we retain one of the independence assumptions but reject the other, we get exactly the right results: black ravens confirm the hypothesis that all ravens are black, but black non-ravens and non-black non-ravens are irrelevant. The assumption retained is that the probability of sampling a raven at random from the universe is independent of the proposition that all ravens are black. This can be explained by the fact that our beliefs about the number of ravens do not depend on our beliefs about their color. The assumption rejected is that the probability of sampling a non-black thing is independent of the proposition that all ravens are black. This is false because our beliefs about the number of non-black things *do* depend in part on our beliefs about whether all ravens are black. For example, if we were to learn that all ravens are black, we should slightly increase our estimate of the number of black things and slightly decrease our estimate of the number of non-black things. Ultimately, these dependence claims are grounded in the fact that the kind *raven* is more natural than the kind *black*. This means that the solution applies to any generalization of the form “All F’s are G,” provided that *F* is more natural than *G*.

2. Formulating the Paradox

The paradox can be stated as follows:²

Nicod's Criterion (NC): A proposition of the form "Object a is both F and G" (that is, $FaGa$) confirms the hypothesis "All F's are G."

Equivalence Condition (EC): If E confirms H1, then E also confirms H2, if H1 and H2 are logically equivalent.

These two plausible principles entail the following:

Paradoxical Conclusion (PC): The proposition "Object a is a non-black non-raven" (that is, $\sim Ba\sim Ra$) confirms the hypothesis "All ravens are black."

Why does this follow? (NC) entails that $\sim Ba\sim Ra$ confirms "All non-black things are non-ravens," which is logically equivalent to "All ravens are black."³ So, by (EC), $\sim Ba\sim Ra$ also confirms "All ravens are black." But this is absurd. The observation of a non-black non-raven is irrelevant to the hypothesis that all ravens are black.

² Many slightly different formulations appear in the literature. The first is in Hempel (1945).

³ This is because $\forall x(R(x) \rightarrow B(x))$ is equivalent to $\forall x(\sim B(x) \rightarrow \sim R(x))$. Some philosophers have attempted to solve the paradox by denying that the English "All ravens are black" is equivalent to the first-order $\forall x(R(x) \rightarrow B(x))$, and therefore denying that the English "All ravens are black" is equivalent to the English "All non-black things are non-ravens." A discussion of this approach is beyond the scope of this paper. Bayesian commentators have typically accepted that these hypotheses are equivalent, and I will do the same.

The first step is to evaluate (NC) and (EC) within a Bayesian framework. However, these principles need to be re-stated because in their current form they fail to mention background knowledge. On Bayesianism, confirmation is a three-place relation between an evidence proposition (E), a hypothesis (H), and background knowledge (K). E confirms H relative to K just in case $P(H|E\&K) > P(H|K)$, which holds if and only if $P(E|H\&K) > P(E|K)$ if and only if $P(E|H\&K) > P(E|\sim H\&K)$. (Probabilities are interpreted as subjective degrees of belief, or credences.) It is easy to see how (EC) should be re-stated:

Equivalence Condition* (EC*): If E confirms H1 relative to background knowledge K, then E also confirms H2 relative to the same background knowledge K, if H1 and H2 are logically equivalent.

(EC*) is true according to Bayesianism, since logically equivalent propositions have the same probability.⁴ However, things are less straightforward in the case of (NC). Here is one version:

Nicod's Criterion, Revision 1 (NCR1): "FaGa" confirms "All F's are G" relative to *all possible* background knowledge K.

⁴ In more detail: Suppose E confirms H1 relative to K. That is, $P(H1|E\&K) > P(H1|K)$. Since H1 and H2 are logically equivalent, $P(H1|J) = P(H2|J)$, for all J. So, $P(H2|E\&K) = P(H1|E\&K)$ and $P(H2|K) = P(H1|K)$. So $P(H2|E\&K) > P(H2|K)$.

In combination with (EC*), (NCR1) yields:

Paradoxical Conclusion, Revision 1 (PCR1): $\sim Ba \sim Ra$ confirms “All ravens are black” relative to all possible background knowledge K.

(PCR1) is truly absurd. However, the Bayesian is not committed to (PCR1), since it is easy to show that (NCR1) is false in a Bayesian framework. Good (1967) provides what is probably the earliest counterexample to (NCR1), but others appear in Swinburne (1971) and Rosenkranz (1982).⁵

One might conclude from this that Bayesianism solves the ravens paradox, by showing that one of its key premises is false. However, I think this would be premature. (NC) has considerable plausibility. Even though (NCR1) is false, there might be some other version of (NC) that is true. If so, this principle might give rise to a problematic version of (PC).

I will describe two other versions of (NC) that are true in a Bayesian framework. However, neither gives rise to a genuinely problematic version of (PC). Here is the first:

⁵ Here's a simple counterexample to (NCR1), similar to the one found in Good 1967: Suppose you know that if all ravens are black, there is only one raven; but if not all ravens are black, there are millions of ravens, most of which are black but one of which is white. Relative to this background knowledge, you are much more likely to encounter a black raven if the ravens hypothesis is false than if it's true; so, the observation of a black raven does not confirm “All ravens are black” (in fact, it disconfirms it).

Nicod's Criterion, Revision 2 (NCR2): Let "a" refer to an object known to have been sampled at random from the set of F's. Suppose we know nothing else about object *a*. Then, FaGa confirms the hypothesis that all F's are G.⁶

Since we know that object *a* was sampled from the set of F's, our background information, K, includes the fact that object *a* is F. However, we know nothing else about object *a*; in particular, we don't know whether object *a* has property G or not. Thus, $P(\text{FaGa}|\text{K}) < 1$. However, $P(\text{FaGa}|\text{All F's are G} \ \& \ \text{K}) = 1$. So, FaGa confirms "All F's are G" relative to this particular K.

(NCR2), in combination with (EC*), entails the following:

"Paradoxical" Conclusion, Revision 2 (PCR2): If we know that object *a* was sampled at random from the set of non-black things, and we know nothing else about object *a*, then $\sim\text{Ba}\sim\text{Ra}$ confirms "All ravens are black."

However, (PCR2) is not genuinely paradoxical. $\sim\text{Ba}\sim\text{Ra}$ *should* confirm "All ravens are black" in this context. Since object *a* was chosen from the set of non-black things, if the ravens hypothesis is true, then it must be a non-raven. However, if the ravens hypothesis is false, it might be a raven. Learning that object *a* is, in fact, not a raven vindicates a prediction of the ravens hypothesis, thereby confirming it. There is no paradox here.

⁶ This version of (NC) plays a large role in treatments of the paradox by Royall (1997), Godfrey-Smith (2003), Horwich (1982), and others.

Here is the second true version of Nicod's Criterion:

Nicod's Criterion, Revision 3 (NCR3): A proposition of the form " $Fa \rightarrow Ga$ " (" \rightarrow " is the material conditional) confirms "All F's are G" relative to any background information K such that $P(Fa \rightarrow Ga|K) < 1$.

$Fa \rightarrow Ga$ simply says that object a is not a non-G F. So, $P(Fa \rightarrow Ga|All\ F's\ are\ G\ \&\ K) = 1$, which means that $P(Fa \rightarrow Ga|All\ F's\ are\ G\ \&\ K) > P(Fa \rightarrow Ga|K)$. So, $Fa \rightarrow Ga$ confirms "All F's are G" relative to K.

(NCR3), in combination with (EC*), yields (PCR3):

"Paradoxical" Conclusion, Revision 3 (PCR3): A proposition of the form " $\sim Ba \rightarrow \sim Ra$ " confirms "All ravens are black" relative to any background information K such that $P(\sim Ba \rightarrow \sim Ra|K) < 1$.

Once again, this version of (PC) is not genuinely paradoxical. $\sim Ba \rightarrow \sim Ra$ is logically equivalent to $Ra \rightarrow Ba$; both say simply that object a is not a non-black raven. Learning *only* that a particular object is not a counterexample to a hypothesis should confirm that hypothesis. This is true regardless of whether the object is in fact a black raven, or a non-black non-raven, or a black non-raven (learning $\sim Ba \rightarrow \sim Ra$ doesn't tell us which).

We have discussed three different precisifications of (NC). The first, most general version *does* give rise to a problematic version of (PC). However, it is false. The second two versions of (NC) are true; but, the versions of (PC) that they entail are not genuinely problematic. I take this to be good evidence that the ravens paradox does not arise in a Bayesian framework.

Even if this is right, though, there is still an unanswered question. Why is it that, upon first hearing the paradox, we find it so natural to think that black ravens confirm the hypothesis (H) that all ravens are black and non-black non-ravens are irrelevant? These particular evidential relations seem privileged in a way that is not explained by what has been said thus far. It's true that finding a known raven to be black confirms H; but, likewise, finding a known non-black object to be a non-raven confirms H. Learning " $Ra \rightarrow Ba$ " confirms H; but, so does " $\sim Ba \rightarrow \sim Ra$." These cases that might loosely be described as confirmation of H by non-black non-ravens are unproblematic; but, the symmetry of these examples does not explain the asymmetry that most people feel upon first hearing the ravens paradox. Does Bayesianism have the resources to explain this?

Many Bayesians think so. Indeed, this is the purpose of the so-called "canonical" Bayesian solution.

3. Existing Bayesian Solutions

The basic idea behind the canonical Bayesian solution first appears in Hosiasson-Lindenbaum (1940). Versions of it are defended in Mackie (1963), Alexander (1958),

Howson and Urbach (1993), and many others. However, as several authors (Horwich (1982), Maher (1999), Vranas (2004), and others) have noted, the canonical solution has a problematic consequence: it entails that black non-ravens disconfirm the hypothesis that all ravens are black.

The aim of the canonical solution, as noted in the previous section, is to find an asymmetry between the evidential import of black ravens and non-black non-ravens. More specifically, the aim is to establish the two results below. For simplicity, we assume that all we know about object a is that it was chosen at random from the (finite) set of all things in the universe.

Result 1: $\sim Ba \sim Ra$ does confirm “All ravens are black,” relative to our actual background knowledge, but the degree of confirmation is minute.

Result 2: $RaBa$ confirms “All ravens are black,” relative to our actual background knowledge, much more strongly than does $\sim Ba \sim Ra$.

So, although $BaRa$ and $\sim Ba \sim Ra$ both confirm H , the degree of confirmation of the former is substantial, while the degree of confirmation of the latter is tiny—so tiny, in fact, that we mistake it for no confirmation at all.

Defenders of the canonical solution show that their two results follow from these three assumptions:

$$A1) P(\sim Ba|K) \gg P(Ra|K)$$

$$A2) P(Ra|H\&K) = P(Ra|K)$$

$$A3) P(\sim Ba|H\&K) = P(\sim Ba|K)$$

A1 is plausible, given that K represents our actual background knowledge. The other two assumptions, A2 and A3, are known as the “independence assumptions,” and they have been criticized in the literature. Both Horwich (1982) and Vranas (2004) argue that, although A3 is not particularly implausible, it is not particularly plausible, either. There doesn't seem to be any good justification for it. Moreover, A2 and A3 jointly entail the problematic consequence mentioned earlier, namely, that $Ba\sim Ra$ disconfirms H. That is, if you choose an object at random from the universe and find that it's a black non-raven, it follows from A2 and A3 that this information must lower your confidence that all ravens are black. But this is absurd; black non-ravens are just as irrelevant to the ravens hypothesis as non-black non-ravens.

We might pause here to reflect on methodology. I have been assuming that any adequate solution to the paradox must accommodate our judgments that $BaRa$ confirms H, and $\sim Ba\sim Ra$ and $Ba\sim Ra$ are neutral. But why am I entitled to assume this, given that

mistakes about basic probabilistic matters are not uncommon? Recall, for example, the base rate fallacy, Simpson's paradox, the gambler's fallacy, the Monty Hall problem, etc.⁷

The moral of these findings, I take it, is that, before endorsing a probabilistic intuition, we should make sure it is not an instance of a known fallacy, and think carefully about whether there is any other compelling theoretical argument against it. The moral is *not* that we should discard every one of our intuitions that concerns probability. Indeed, this would spell disaster, as we need to rely on probabilistic judgments all the time in ordinary life. For example, I judge that a green light is evidence that it is safe to cross the intersection; I judge the presence of a breeze to be neutral (no evidence either way) with respect to the location of my car in a parking lot; I judge that a helicopter's looking red is evidence that it is red; etc. In a Bayesian framework, every one of these is a probabilistic judgment, just like the judgments that black ravens are evidence that all ravens are black, but non-black non-ravens and black non-ravens are neutral. We are entitled to rely on such judgments, so long as they are not an instance of any known fallacy, and there is no other compelling theoretical argument against them. The three judgments I presuppose in this paper pass both tests. Thus, it is a serious problem that A2 and A3—two assumptions for which no justification has been provided—have the counterintuitive consequence that black non-ravens disconfirm “All ravens are black.”

Why do A2 and A3 have this consequence? Understanding this will be helpful later on, when I present my improved Bayesian solution. First, consider these four mutually

⁷ Thanks to an anonymous reviewer for raising this point.

exclusive and exhaustive propositions: $BaRa$, $Ba\sim Ra$, $\sim BaRa$, and $\sim Ba\sim Ra$. Initially you have some non-zero credence in each, as depicted in table 1 below. $P(BaRa) = q$, $P(Ba\sim Ra) = r$, etc.⁸ In accordance with the laws of probability, $q + r + s + t = 1$.

Table 1: Initial Credences

	R	$\sim R$
B	q	r
$\sim B$	s	t

Now consider your credences conditional on H. $P(\sim BaRa|H) = 0$; so, your credences for the other three propositions must now sum to 1. The credence you initially had in $\sim BaRa$ (s in table 1) must be distributed among these remaining three possibilities. The crucial question is, how?

⁸ For simplicity I have omitted “K,” but the reader should keep in mind that credences are always implicitly conditional on your background knowledge K.

Table 2: Credences Conditional on H

	R	$\sim R$
B	?	?
$\sim B$	0	?

A2 says that $P(Ra|H) = P(Ra)$; equivalently, $P(BaRa|H) + P(\sim BaRa|H) = P(BaRa) + P(\sim BaRa)$. That is, the sum of the values in the left-hand column (under R) in table 2 must equal the sum of the values in the left-hand column in table 1, namely, $q + s$. Since $P(\sim BaRa|H) = 0$, it follows that $P(BaRa|H) = q + s$, as in table 3:

Table 3: Credences Conditional on H, With A2

	R	$\sim R$
B	$q + s$?
$\sim B$	0	?

This makes it look like the effect of conditionalizing on H is to redistribute your initial credence for $\sim BaRa$ (namely, s) by adding it to your credence for $BaRa$. But see what happens when we add the A3 assumption: $P(\sim Ba|H) = P(\sim Ba)$, equivalently, $P(\sim Ba\sim Ra|H) +$

$P(\sim\text{BaRa}|H) = P(\sim\text{Ba}\sim\text{Ra}) + P(\sim\text{BaRa})$. A3 requires that the sum of the values in the bottom row must equal the sum of the values in the bottom row in table 1, namely, $s + t$. Since $P(\sim\text{BaRa}|H) = 0$, it follows that $P(\sim\text{Ba}\sim\text{Ra}|H) = s + t$, as in table 4:

Table 4: Credences Conditional on H, With A2 and A3:

	R	$\sim R$
B	$q + s$?
$\sim B$	0	$s + t$

This is a bit puzzling. It looks as though A2 tells you to redistribute your old credence for $\sim\text{BaRa}$ (namely, s) by adding it all to BaRa ; but, A3 seems to tell you to redistribute by adding it all to $\sim\text{Ba}\sim\text{Ra}$. How can you do both? You only have s amount of newly-freed credence, not $2s$.

The answer is inescapable: the only way to satisfy both A2 and A3 is to subtract precisely s from your initial credence in the fourth proposition, $\text{Ba}\sim\text{Ra}$, as in table 5:

Table 5: Credences Conditional on H, With A2 and A3:

	R	$\sim R$
B	$q + s$	$r - s$
$\sim B$	0	$s + t$

Accepting both A2 and A3 forces us to set $P(\text{Ba}\sim\text{Ra}|\text{H}) = P(\text{Ba}\sim\text{Ra}) - s$.⁹ That is why these two assumptions entail that $\text{Ba}\sim\text{Ra}$ disconfirms H.

The canonical Bayesian solution is in trouble. Its reliance on the combination of A2 and A3 results in the consequence that black non-ravens disconfirm “All ravens are black.” This is just as problematic as the consequence the canonical solution was designed to address—namely, that non-black non-ravens confirm “All ravens are black.” The canonical solution has simply traded one problem for another.¹⁰

In the next section I present an improved Bayesian solution that involves accepting A2 but rejecting A3. On this solution, credences conditional on H are as displayed in table 6.

⁹ Here is a more formal version of the proof. The laws of probability dictate the following: $P(\text{BaRa}) + P(\text{Ba}\sim\text{Ra}) + P(\sim\text{BaRa}) + P(\sim\text{Ba}\sim\text{Ra}) = 1 = P(\text{BaRa}|\text{H}) + P(\text{Ba}\sim\text{Ra}|\text{H}) + P(\sim\text{BaRa}|\text{H}) + P(\sim\text{Ba}\sim\text{Ra}|\text{H})$. A2 says: $P(\text{BaRa}|\text{H}) + P(\sim\text{BaRa}|\text{H}) = P(\text{BaRa}) + P(\sim\text{BaRa})$. A3 says: $P(\sim\text{Ba}\sim\text{Ra}|\text{H}) + P(\sim\text{BaRa}|\text{H}) = P(\sim\text{Ba}\sim\text{Ra}) + P(\sim\text{BaRa})$. Replacing $P(\sim\text{BaRa})$ with s and $P(\sim\text{BaRa}|\text{H})$ with 0 yields: $P(\text{BaRa}) + P(\text{Ba}\sim\text{Ra}) + s + P(\sim\text{Ba}\sim\text{Ra}) = P(\text{BaRa}|\text{H}) + P(\text{Ba}\sim\text{Ra}|\text{H}) + 0 + P(\sim\text{Ba}\sim\text{Ra}|\text{H})$, and A2: $P(\text{BaRa}|\text{H}) = P(\text{BaRa}) + s$, and A3: $P(\sim\text{Ba}\sim\text{Ra}|\text{H}) = P(\sim\text{Ba}\sim\text{Ra}) + s$. Substituting the A2 and A3 consequences into the equality yields: $P(\text{BaRa}) + P(\text{Ba}\sim\text{Ra}) + s + P(\sim\text{Ba}\sim\text{Ra}) = P(\text{BaRa}) + s + P(\text{Ba}\sim\text{Ra}|\text{H}) + P(\sim\text{Ba}\sim\text{Ra}) + s$. Many terms cancel to yield $P(\text{Ba}\sim\text{Ra}|\text{H}) = P(\text{Ba}\sim\text{Ra}) - s$.

¹⁰ Note that assumptions A1 – A3 do *not* entail that the degree of disconfirmation by $\text{Ba}\sim\text{Ra}$ will be minute.

Table 6: Credences Conditional on H, According to the New Solution:

	R	$\sim R$
B	$q + s$	r
$\sim B$	0	t

This way of redistributing credence yields the desired results exactly: our credences for $Ba\sim Ra$ and $\sim Ba\sim Ra$ are unchanged, and so these two pieces of evidence are neutral. However, $P(BaRa|H) > P(BaRa)$, and so $BaRa$ confirms. In filling out the solution, I will provide a deeper reason that explains why our credence should be re-distributed in this way, and why A2 but not A3 should be accepted.

First, though, I will discuss another existing Bayesian treatment of the paradox, in Fitelson and Hawthorne (2010). They aim to establish a merely comparative result about the strength of confirmation of $BaRa$ versus $\sim Ba\sim Ra$. More specifically, let $c(E,H|K)$ measure the degree to which E confirms H relative to background knowledge K . Fitelson and Hawthorne seek to establish that $c(BaRa,H|K) > c(\sim Ba\sim Ra,H|K)$. They show that this result follows from these two assumptions:

FHA1) $P(\sim Ba|K) \gg P(Ra|K)$

FHA2) $P(\sim Ba|K) / P(Ra|K)$ is greater than or approximately equal to¹¹

$$P(\sim Ba|H\&K) / P(Ra|H\&K)$$

FHA1 is the same as A1 above. FHA2 is strictly weaker than A2 and A3; it is entailed by them. Fitelson and Hawthorne do not aim to provide a justification for it; they argue only that it is *more plausible* than A2 and A3, since it is logically weaker. Moreover, it is superior to the A2/A3 combination because it does not entail the problematic result that $Ba\sim Ra$ disconfirms H. Although Fitelson and Hawthorne do not provide a justification for it, my considered view is that FHA2 should be accepted, because it is entailed by the solution I develop in the next section.

So, I think that Fitelson and Hawthorne are on the right track, but that more can be said. Since their result is merely comparative, it is compatible with strong disconfirmation by both $BaRa$ and $\sim Ba\sim Ra$ (in which case $BaRa$ would disconfirm less) and also compatible with strong confirmation by both (in which case $BaRa$ would confirm more). So while their result does establish an asymmetry between these two pieces of evidence, it doesn't explain why $BaRa$ *confirms* and why $\sim Ba\sim Ra$ is *neutral*. The solution I present in the next section entails these stronger results, and is entirely compatible with Fitelson and Hawthorne's approach.

¹¹ The precise meaning of this is explained in their technical addendum: http://fitelson.org/wason_addendum.pdf.

Let's take stock. Although we did not see any way to formulate a version of the traditional ravens paradox in the Bayesian framework, the following question remained: how do we explain the persistent thought that certain evidential relations—that $BaRa$ confirms, and $\sim Ba\sim Ra$ is neutral—are privileged? We considered the most popular Bayesian attempt to answer this question (the canonical solution), and found that it has a serious problem: when combined, the A2 and A3 assumptions yield the problematic consequence that $Ba\sim Ra$ disconfirms. We considered a recent Bayesian approach that does not appeal to A2 and A3 (Fitelson and Hawthorne (2010)), and found that, though it sought to establish an asymmetry between black ravens and non-black non-ravens, it didn't explain the intuition that black ravens should *confirm*, and non-black non-ravens should be *neutral*.

4. A New Bayesian Solution

Recall the problem at hand. Object a was chosen at random from the universe. Although we know that there is some possible background information relative to which $BaRa$ disconfirms H , and some possible background information relative to which $\sim Ba\sim Ra$ significantly confirms H , the thought persists that certain evidential relations are privileged: that $BaRa$ confirms H , and $\sim Ba\sim Ra$ and $Ba\sim Ra$ are irrelevant to it. How can this be explained?

Recall from the previous section that the evidential import of these types of evidence is determined by the way in which our credence for $\sim BaRa$ is re-distributed upon conditionalizing on H . Conditional on H , our credence in $\sim BaRa$ goes to zero, and we must

distribute this credence somehow among the remaining three possibilities: $BaRa$, $\sim Ba\sim Ra$, and $Ba\sim Ra$.

The crucial thing to notice is that, as noted in the previous section, one particular way of redistributing credence yields exactly the desired results. Suppose the only effect of conditionalization on H is that our pre- H credence in $\sim BaRa$ shifts over to be added to our credence in $RaBa$. Then our pre- and post- H credences in $\sim Ba\sim Ra$ and $Ba\sim Ra$ are unaffected; and so, these pieces of information are neutral to H . We needn't try to explain away our intuition that $\sim Ba\sim Ra$ is neutral by claiming that we mistake minuscule confirmation for no confirmation—rather, our intuition is that $\sim Ba\sim Ra$ should be neutral because it is, as a matter of fact, neutral. Our pre- H credence in $BaRa$ is less than our post- H credence in it, and so $BaRa$ confirms H . Here, copied from the previous section, are two tables that depict this particular way of redistributing credence:

Table 1: Initial Credences

	R	$\sim R$
B	q	r
$\sim B$	s	t

Table 6: Credences Conditional on H, According to the New Solution:

	R	$\sim R$
B	q + s	r
$\sim B$	0	t

This shows that Bayesianism is compatible with our intuitions. But there is a further question: Why is it that our credence should move in this way? Although we have seen that it yields intuitive results, it is natural to wonder whether there is a deeper reason. At the very least, there is one puzzle that needs to be cleared up: redistributing credence in this way commits us to accepting A2 but denying A3. In a sense this is good, because we saw that accepting both leads to trouble. But in a sense it is puzzling, because we wonder what the

relevant asymmetry is. What is wrong with A3, and why is it that we should retain A2 at the expense of A3? Is there a deeper reason for this?

Indeed, there is. First, notice that the following thought would naturally motivate A2: suppose that your beliefs concerning the number of ravens are unaffected by learning H. If so, then it would be natural to adopt the A2 assumption, according to which the probability of sampling a raven from the universe is unchanged by conditionalization on H.¹² Alternatively, however, suppose your beliefs about the number of black things were unchanged by learning H. This would naturally motivate the A3 assumption, on which the probability of sampling a black thing is unchanged by conditionalization on H.¹³

The question, then, is this: does it make sense to think that your beliefs about the number of ravens are unchanged by learning H, but that your beliefs about the number of black things would be changed by learning H? If so, what is the asymmetry?

The asymmetry is that beliefs about the number of black things depend in part on beliefs about the color of ravens—such as how many are black—whereas beliefs about the number of ravens do not depend on beliefs about their color. So, learning that all ravens are black changes your beliefs about the number of black things, but does not affect your beliefs about the number of ravens.¹⁴

Let's begin with the number of black things. If you were to try to estimate this number, it would be natural to begin by considering which types of things have some black

¹² Some readers may notice that this depends, also, on my beliefs about the size of the universe. This is discussed below.

¹³ Above the A3 assumption is stated as $P(\sim Ba|H) = P(\sim Ba)$; this is equivalent to $P(Ba|H) = P(Ba)$.

¹⁴ There are some possible cases in which this fails, as I explain later.

instances (for example, chairs, mushrooms, ravens, etc.), and then considering, for each of these types, how many are black. (How many of the chairs are black? How many of the ravens? etc.) This reveals that our beliefs about the number of black things depend in part on our beliefs about the color of ravens. Prior to learning H we think that some of the ravens might be non-black, and so it may be that only a fraction of the ravens contribute to the total number of black things. However, learning H tells us that *all* of the ravens are black, and so *all* of the ravens contribute to the total number of black things. Since this information doesn't affect our beliefs about the numbers of other kinds of things (such as chairs or mushrooms) or the percentages of them that are black, overall one's estimate of the number of black things should go up upon learning H.

Now consider our beliefs about the number of ravens. It would be very *unnatural* to attempt an estimate of the number of ravens by trying to answer questions like these: First, how many black things are there? Now, what percentage of them are ravens? Next, how many white things are there, and what percentage of *them* are ravens? How many red things are there, and what percentage are ravens? Etc. This reveals that our beliefs about the number of ravens do *not* depend on beliefs of this kind. If they did, then learning H would affect our beliefs about the number of ravens. Learning H tells us that the percentage of white things that are ravens is zero, the percentage of red things that are ravens is zero, etc. So, if it *did* make sense to estimate the number of ravens in this way, then learning H *would* affect our beliefs about the number of ravens, via its effect on our beliefs about the percentages of white, red, etc. things that are ravens. However, it would be extremely

unnatural to estimate the number of ravens in this way. Our beliefs about the number of ravens do not depend on beliefs of this kind, and so they are unaffected by learning H.

Of course, it is true, as a bare mathematical fact, that the number of ravens is equal to the number of black things that are ravens plus the number of white things that are ravens plus...etc. for all the colors. What I am denying is that our beliefs about the former *depend* on our beliefs about the latter, in the distinctively epistemic sense of dependence that I am employing here.

This epistemic dependence relation is closely connected with the basing relation. In general, if an agent's belief about whether P is based (at least in part) on the agent's belief about whether Q—and the agent is not irrational in so basing her belief—then whether P depends, epistemically, on whether Q. I will not attempt to give necessary and sufficient conditions for dependence; the notion is intuitive enough that we can apply it to the cases discussed in this paper without a precise definition. However, I will give a few examples, to make sure we all have the same concept in mind.

Relations of epistemic dependence are commonplace. One's belief about what the temperature is depends on one's belief about what the thermometer reads. The detective's belief about whether the butler did it depends (in part) on his belief about whether the butler's fingerprints were found at the scene of the crime. And, I claim, our beliefs about the number of black things depend, in part, on our beliefs about whether all ravens are black, while our beliefs about the number of ravens do *not* depend on our beliefs about whether all ravens are black.

Is there a deeper reason why these epistemic dependence relations hold? Why does it seem so natural to estimate the number of black things by estimating the number of ravens, and other kinds, and the percentages of each that are black, but so *unnatural* to estimate the number of ravens by estimating the number of black things, and the other colors, and the percentages of each that are ravens?

The answer, I suggest, is that *raven*, the natural kind, is *more natural* than the kind *black*. This is what ultimately explains why it makes sense to estimate the number of *black things* via a procedure that involves holding fixed the number of *ravens*, but it does *not* make sense to estimate the number of *ravens* via a procedure that involves holding fixed the number of *black things*.

Naturalness also plays a role in the famous solution to the ravens paradox proposed in Quine 1969, but in a way that is very different from the solution presented here. Quine suggested restricting Nicod's Criterion to natural kinds. The predicates *raven* and *black* are both considered natural, so this proposal correctly predicts that black ravens should confirm "All ravens are black." However, *non-black* and *non-raven* do not count as natural, which allows the proposal to explain why non-black non-ravens do not confirm "All non-black things are non-ravens" (which, recall, is equivalent to "All ravens are black.") However, Quine's restriction is too severe. There are some generalizations involving non-natural predicates that do seem instance-confirmable (for example, "All ravens are non-white.")

Quine was right that naturalness is a key part of the solution to the ravens paradox. However, the proper role for naturalness is different from the one he proposed. We should

not issue an absolute prohibition against predicates that are not deemed sufficiently natural. Rather, the crucial issue concerns *comparative degrees* of naturalness, and, in particular, whether F is more natural than G in a generalization of the form “All F 's are G 's.”

Here, then, is an overview of my proposal. That *raven* is more natural than *black* grounds the fact that our beliefs about the number of ravens *do not* depend on our beliefs about the color of ravens, but our beliefs about the number of black things *do* depend (in part) on our beliefs about the color of ravens. These dependence relations, in turn, explain why we should adopt the A2 assumption but reject the A3 assumption, which explains why $BaRa$ confirms H but $\sim Ba\sim Ra$ and $Ba\sim Ra$ are neutral. Below I will provide a more formal presentation this line of thought, but first I need to address a potential worry.

The view just described succeeds in explaining why we find certain evidential relations so intuitively plausible—that $BaRa$ confirms H , while $\sim Ba\sim Ra$ and $Ba\sim Ra$ are neutral. However, we know that by changing the background information, we can get the evidential relations to come out differently. One might worry that the view just described cannot accommodate this fact. After all, it was claimed that what ultimately grounds the privileged evidential relations is that *raven* is more natural than *black*. But this difference in naturalness holds regardless of what background information one has. How, then, can the evidential relations be appropriately sensitive to background information?

The answer is that certain kinds of background information can sever the connection that otherwise would hold between the naturalness facts and the dependence facts.

It will help to begin by considering cases in which one's background information is very minimal—one knows only that some ravens exist, and has whatever empirical knowledge (if any) is required to grasp ordinary concepts such as *raven* and *black*.¹⁵ Since there is no background information that might prevent it from doing so, the difference in naturalness inherent in the predicates *raven* and *black* will determine the dependence relations, as described above: beliefs about the number of ravens will not depend on beliefs about their color, but beliefs about the number of black things will.

Moreover, there is a very wide range of different background information one could acquire, including our actual background information, that would leave this connection untouched. We actually know quite a lot about ravens—we know that they are particularly intelligent birds, that they are larger than crows, that they are widespread in the United States, etc. But none of this affects the dependence relations, and so they are still determined by the difference in naturalness between *raven* and *black*, just as in the case of minimal background information.

However, some possible background information *would* get in the way of this connection. For example, suppose you know that there are a certain number of ravens living in the temperate zones of the world, all of which are black. You also know that there is a certain population of white birds in the Arctic, but you do not know whether they are crows

¹⁵ Some authors (including Good (1968) and Maher (2004)) have argued that relative to *merely tautological* background information, finding a black raven could actually *disconfirm* “All ravens are black.” Their claim is highly controversial; space constraints prevent me from discussing the debate here. However, their argument does not go through if we know that there are at least some ravens, as I have stipulated. So it does not endanger my claim that black ravens confirm relative to this *minimal* background information.

or ravens. There are no other possible raven populations. Relative to this background knowledge, learning H *would* affect your beliefs about the number of ravens. Given what you know, H is true if and only if the white Arctic birds are crows rather than ravens. So, if H is true, the white Arctic birds are not ravens, and so your estimate of the total number of ravens would go down. But learning H would *not* affect your beliefs about the number of black things, because you already know which ravens contribute to the number of black things, regardless of whether H is true—namely, the ravens that inhabit the temperate zones.¹⁶

As we might expect, since this background information undercuts the connection that would otherwise hold between the naturalness facts and the dependence relations, the evidential relations are also altered. Now, sampling a black raven from the universe provides no evidence whatsoever for H, because the number of *black* ravens is the same, regardless of whether H is true (i.e. regardless of whether the white Arctic birds are ravens or crows). However, finding a non-black non-raven confirms H, because there will be more of them if H is true (because then the white Arctic birds are crows, and so contribute to the overall number of non-black non-ravens).

The upshot is that the solution proposed here has the flexibility to accommodate differences in evidential relations due to differences in background information, while nevertheless explaining the persistent thought that certain evidential relations are privileged. Relative to minimal background information, the difference in degree of naturalness between

¹⁶ Thanks to Andrew Graham and an anonymous reviewer, who independently suggested that I discuss cases of this kind.

raven and *black* makes it the case that beliefs about the number of black things do, but beliefs about the number of ravens do not, depend on beliefs about the color of ravens. This, in turn, helps explain why BaRa confirms H, while \sim Ba \sim Ra and Ba \sim Ra are neutral. Moreover, a very wide range of possible background information, including our actual information, leaves these connections undisturbed. In such cases, these privileged epistemic relations hold in virtue of properties inherent to the predicates themselves: the fact that the kind *raven* is more natural than the kind *black*.

Henceforth I will presuppose that we are focusing on cases in which this connection is undisturbed, but it should not be forgotten that there is always some possible background information that would undercut this effect of the difference in naturalness.

Returning to the main thread, there is one extra complexity that needs to be taken into account. I have claimed that learning H affects our beliefs about the number of black things, but not the number of ravens, and that this explains why we should accept A2 but reject A3. However, the probability of sampling a raven from the universe depends not only on our beliefs about the number of ravens, but also our beliefs about the size of the universe. There is, then, a gap in the informal argument I've been developing for A2 and against A3.

However, the gap is easily filled, because our beliefs about the number of things in the universe are unaffected by learning H. This is because these beliefs depend on beliefs of the following kind: how many chairs there are; how many trees; how many frogs; how many ravens; etc. for all the kinds that are instantiated. Since one's estimate of none of these is

affected by learning what color ravens are, one's beliefs about the number of things in the universe is not affected by learning H.

It's time to give a more formal presentation of the line of thought I've been developing. What I aim to establish, recall, is that conditionalization on H should prompt only the following change: that we redistribute our old probability for \sim BaRa by adding it to our credence for BaRa. This is encapsulated in the following three desired conclusions:

$$(C1) P(\text{BaRa}|\text{H}) = P(\text{BaRa}) + P(\sim\text{BaRa})$$

$$(C2) P(\text{Ba}\sim\text{Ra}|\text{H}) = P(\text{Ba}\sim\text{Ra})$$

$$(C3) P(\sim\text{Ba}\sim\text{Ra}|\text{H}) = P(\sim\text{Ba}\sim\text{Ra})$$

Below, I will prove that these conclusions follow from premises P1 – P3:

$$P1) \forall i \forall k [P(\text{R} = i \ \& \ \text{U} = k) = P(\text{R} = i \ \& \ \text{U} = k|\text{H})]^{17}$$

$$P2) \forall i \forall k [P(\text{BNR} = i \ \& \ \text{U} = k) = P(\text{BNR} = i \ \& \ \text{U} = k|\text{H})]$$

$$P3) \forall i \forall k [P(\text{NBNR} = i \ \& \ \text{U} = k) = P(\text{NBNR} = i \ \& \ \text{U} = k|\text{H})]$$

Notation is as follows: R = the number of ravens; U = the number of things in the universe; BNR = the number of black non-ravens; NBNR = the number of non-black non-ravens. Premises P1 – P3 are justified by claims about epistemic dependence of the sort

¹⁷ I assume that there is some large number n such that for all $k > n$, $P(\text{U} = k) = 0$.

described above. Let's begin with P1. Above, I noted that our beliefs about the number of ravens and the number of things in the universe are unaffected by learning H. We can express this more precisely by saying that our credences in propositions of the form $R = i$ & $U = k$ are unaffected by conditionalization on H, which is precisely what P1 says.

Turn now to P2. We have seen that our beliefs about the number of things in the universe are unaffected by H, but what about our beliefs concerning the number of non-ravens that are black? These beliefs depend on answers to questions about what types there are other than ravens (e.g. chairs, frogs, trees, etc.) and, for each type, what percentage are black. Such beliefs are not affected by H, which is a claim about the color of *ravens*, and does not affect our beliefs about the colors of types other than ravens (for example, how many of the shoes are black).

Similar considerations motivate P3: our beliefs about the number of non-black non-ravens depend, again, on our beliefs about which types of things there are, other than ravens, and what percentage of each type is non-black. Once again, learning H—a claim about the color of ravens—doesn't affect our beliefs about the color of non-ravens, such as tables, chairs, or bowls.

It remains only for me to prove that the desired conclusions C1 – C3 do, indeed, follow from the premises P1 – P3:

By the laws of probability (specifically, the theorem of total probability):

$$(1) P(\text{Ra}) = \sum_i \sum_k [P(\text{Ra}|\text{R} = i \ \& \ \text{U} = k)P(\text{R} = i \ \& \ \text{U} = k)]^{18}$$

The Principal Principle (Lewis (1986)) lets us replace $P(\text{Ra}|\text{R} = i \ \& \ \text{U} = k)$ with (i / k) :

$$(2) P(\text{Ra}) = \sum_i \sum_k [(i / k)P(\text{R} = i \ \& \ \text{U} = k)]$$

By the laws of probability (again, the total probability theorem):

$$(3) P(\text{Ra}|H) = \sum_i \sum_k [P(\text{Ra}|\text{R} = i \ \& \ \text{U} = k \ \& \ H)P(\text{R} = i \ \& \ \text{U} = k|H)]$$

The Principal Principle lets us replace $P(\text{Ra}|\text{R} = i \ \& \ \text{U} = k \ \& \ H)$ with (i / k) :

$$(4) P(\text{Ra}|H) = \sum_i \sum_k [(i / k)P(\text{R} = i \ \& \ \text{U} = k|H)]$$

Our first premise, P1, says: $\forall i \ \forall k [P(\text{R} = i \ \& \ \text{U} = k) = P(\text{R} = i \ \& \ \text{U} = k|H)]$

From (2), (4), and P1 we infer $P(\text{Ra}|H) = P(\text{Ra})$; therefore,

$$P(\text{BaRa}|H) + P(\sim\text{BaRa}|H) = P(\text{BaRa}) + P(\sim\text{BaRa})$$

Since $P(\sim\text{BaRa}|H) = 0$, we have our first desired conclusion:

$$(C1) P(\text{BaRa}|H) = P(\text{BaRa}) + P(\sim\text{BaRa})$$

Analogous lines of reasoning take us from P2 to C2 and from P3 to C3.

I will now consider an objection. According to my view, if one chooses an object at random from the set of black things and obtains a raven, this constitutes evidence for H. But many authors (including Hempel (1945), Godfrey-Smith (2003), and Royall (1997)) have held that this type of evidence should be neutral, since the procedure of sampling from the black things does not allow for the possibility of turning up a counterexample to H; there is no way that one could falsify H by following this procedure.

¹⁸ I assume there is some large number n such that for all $j > n$, $P(\text{U} = j) = 0$ and $P(\text{R} = j) = 0$.

It is true that one will never find a non-black raven by sampling from the set of black things. However, the mere fact that a procedure does not allow for the possibility of falsification does not mean that there is no possibility for probabilistic confirmation. Once again, this consequence of my view can be explained by plausible assumptions about relations of epistemic dependence. If all ravens are black, then it is reasonable to think that there are more black ravens than one previously thought. However, learning H does not affect our beliefs about the numbers of other kinds of things that are black (for example, the number of black mushrooms). So, we are more likely to find a raven among the black things if H is true. Thus, doing so provides some positive evidence for H .

Finally, I will close with a comment on the generality of the solution. I have focused on the particular example “All ravens are black.” But the core insight of the solution applies much more broadly. Whenever F is more natural than G in a generalization “All F ’s are G ,” we can use the account developed here to explain why certain evidential relations—that $FaGa$ confirms while $\sim Ga\sim Fa$ and $Ga\sim Fa$ are neutral—are privileged. Just as in the ravens, our beliefs about the number of G ’s, but not our beliefs about the number of F ’s, will depend on our belief about whether all F ’s are G ’s, and these dependence relations will ground the privileged evidential relations. F is more natural than G in a great many hypotheses of interest: “All sodium salts burn yellow” (Hempel’s other favourite generalization), “All electrons have negative charge,” “All emeralds are green,” and so forth.

One of the broad morals of the existing literature on the paradox is that evidential relations are highly dependent on the details of one’s background knowledge. The new

analysis given here shows that we can say something more: despite this sensitivity to background information, when F is more natural than G , certain evidential relations are privileged, in the sense that they hold relative to minimal background information, and continue to hold for the same reasons relative to a wide variety of other background information one could acquire. The evidential relations that are privileged in this way are that $FaGa$ confirms “All F ’s are G ’s,” while $GaFa$ and $Ga\sim Fa$ are neutral.

5. Conclusion

It is a weakness of the canonical Bayesian solution to the paradox of the ravens that it requires *both* that the probability of sampling a raven, *and* that the probability of sampling a black thing, be independent of the hypothesis that all ravens are black. Together, these assumptions have the unwelcome consequence that black non-ravens disconfirm the hypothesis. I have developed a new Bayesian solution that involves retaining the former independence assumption but dropping the latter. That it makes sense to do so is explained by appeal to certain epistemic dependence claims. These, in turn, have their source in the fact that the kind *raven* has a greater degree of naturalness than the kind *black*. This approach to the paradox yields exactly the right results: black ravens confirm the hypothesis that all ravens are black, while non-black non-ravens and black non-ravens are irrelevant to it. Moreover, the solution applies to any universal generalization of the form “All F ’s are G ” where F is more natural than G .

References:

- Alexander, H. G. (1958), "The Paradoxes of Confirmation", *British Journal for the Philosophy of Science* 9: 227–233.
- Fitelson, Branden and James Hawthorne (2010), "The Wason Task(s) and the Paradox of Confirmation", *Philosophical Perspectives* 24(1): 207–241.
- Godfrey-Smith, Peter (2003), "Goodman's Problem and Scientific Methodology", *Journal of Philosophy* 100: 573–590.
- Good, Irving John (1967), "The White Shoe is a Red Herring", *British Journal for the Philosophy of Science* 17(4): 322.
- (1968), "The White Shoe qua Herring is Pink", *British Journal for the Philosophy of Science* 19: 156–157.
- Hempel, Carl Gustav (1945), "Studies in the Logic of Confirmation", *Mind* 54: 1–26, 97–121.
- Horwich, Paul (1982), *Probability and Evidence*. New York: Cambridge University Press.
- Hosiasson-Lindenbaum, Janina (1940), "On Confirmation", *The Journal of Symbolic Logic* 5(4): 133–148.
- Howson, Colin and Peter Urbach (1993), *Scientific Reasoning: The Bayesian Approach*. Chicago: Open Court.
- Lewis, David (1986), "A Subjectivist's Guide to Objective Chance", in *Philosophical Papers: Volume II*. Oxford: Oxford University Press, 83–114.

- Mackie, John (1963), "The Paradox of Confirmation", *British Journal for the Philosophy of Science* 13: 265–277.
- Maher, Patrick (1999), "Inductive Logic and the Ravens Paradox", *Philosophy of Science* 66: 50–70.
- (2004), "Probability Captures the Logic of Scientific Confirmation", in Christopher Hitchcock (ed.), *Contemporary Debates in the Philosophy of Science*. Malden: Blackwell, 69–93.
- Quine, Willard Van Orman (1969), "Natural Kinds", in *Ontological Relativity and Other Essays*. New York: Columbia University Press, 114–139.
- Rosenkrantz, Roger (1982), "Does the Philosophy of Induction Rest on a Mistake?", *Journal of Philosophy* 79: 78–97.
- Royall, Richard (1997), *Statistical Evidence: A Likelihood Paradigm*. New York: Chapman & Hall.
- Swinburne, Richard (1971), "The Paradoxes of Confirmation: A Survey", *American Philosophical Quarterly* 8: 318–330.
- Vranas, Peter (2004), "Hempel's Raven Paradox: A Lacuna in the Standard Bayesian Solution", *British Journal for the Philosophy of Science* 55: 545–560.