Hempel's Raven Paradox: On Confirmation and Infinities

Yunus Prasetya
NUS College, National University of Singapore

Hempel's paradox of confirmation can be formulated as the following argument:

> *Nicod's Condition* (NC): For any object *x* and any properties *F* and *G*, the proposition that *x* has both *F* and *G* confirms the hypothesis that every *F* has *G* (Nicod 219).
>
> *Equivalence Condition* (EC): If hypotheses *H1* and H*2* are logically equivalent, then any evidence *E* that confirms *H1* also confirms *H2*.

Together, these two principles entail the following conclusion:

> *Paradoxical Conclusion* (PC): an object *a,* which is a non-black non-raven (*~Ba & ~Ra)* confirms the hypothesis (*H*) that all ravens are black.[1]

This argument is valid. From NC, it follows that *~Ba & ~Ra* confirms the hypothesis that all non-black things are non-ravens. This hypothesis is the contrapositive of *H* and therefore logically equivalent to *H*. So by EC, *~Ba & ~Ra* confirms *H*. But this seems counterintuitive—a white shoe seems irrelevant to whether all ravens are black. Thus, we have an argument that starts from apparently obvious premises and proceeds through valid reasoning to an apparently unacceptable conclusion—a paradox.

A solution to Hempel's paradox should be sufficiently general. Hempel's paradox is about confirmation, not ravens (Clarke 427). It uses ravens only to make the paradox more vivid. Thus, a good solution to the paradox would be applicable not only to black ravens, but also to white swans, cubed dice, and so on. Further, a solution must give us an answer by telling us whether to accept PC, reject NC, or reject EC. But it should also give us an explanation apart from the argument itself for why we should accept its answer. It would not do to merely say that we should accept PC because NC and EC are true. The purported solution should give us a deeper explanation for why we should accept PC.

My aim in this paper is to offer a revision of Nicod's Condition and, in so doing, offer a solution to the raven paradox. First, I explore how Bayesians have observed that NC is insensitive to the role of background information in confirmation. Next I explore a revision of NC that Bayesians have proposed, as well as the standard Bayesian solution to the raven paradox.[2] I argue, however, that the Bayesian revision of NC is still not stringent enough; it fails if there is an infinite number of unobserved *F*'s. In its place, I revise NC by restricting it to apply only to finite classes. Thus, my answer to Hempel's paradox is to accept PC by defending a qualified version of NC. My defense of this qualified NC will also function as a deeper explanation for why PC is true.

I. Nicod's Condition and Background Information

According to Bayesians, confirmation is a three-part relation between a hypothesis, a set of evidence, and background information. If they are right, then NC is flawed for it says nothing about background information. Indeed, it has been shown that NC cannot be true relative to all possible background information (Good 322).[3] Due to this flaw, Bayesians have offered revised versions of NC (e.g. Rinard 84, and Horwich ch. 3). For now, we will use the following revision:

> *Bayesian Nicod's Condition* (BNC): For any randomly sampled object *x* and any properties *F* and *G*, if we know nothing else about *x,* the proposition that *x* has both *F* and *G* confirms the hypothesis that every *F* has *G* (Rinard 84).

BNC is more plausible than NC because it is more stringent. The additional qualifiers—that the object be randomly sampled, and we know nothing else about it—are meant to rule out gerrymandered pieces of information in our background knowledge that might prevent confirmation. In other words, if we have no other special information about the object, then its having *F* and *G* confirms the hypothesis that all *F*'s are *G*'s. In my view, BNC is not sufficiently stringent because though it blocks off pieces of information about *x*, it says nothing about what pieces of information we can have about the world. In particular, I will argue that BNC does not hold if there is an infinite number of unobserved *F*s in the set from whence we are sampling.

II. THE STANDARD BAYESIAN SOLUTION TO THE RAVEN PARADOX

Proponents of the standard Bayesian solution (e.g. Alexander 233, and Mackie 275-76) argue that a non-black non-raven confirms *H*. But they also claim that the degree of confirmation is minute. A non-black non-raven confirms *H* to a much smaller degree than a black raven would. This is because there are many more non-black objects than ravens. Consider the following application of Bayes' theorem on the effect of observing *~Ba & ~Ra* on *H*:

$$P(H|{\sim}Ba \,\&\, {\sim}Ra) = \frac{P({\sim}Ba \,\&\, {\sim}Ra|H)P(H)}{P({\sim}Ba \,\&\, {\sim}Ra)}.$$

From this it can be proven that:

$$P\big(H \,\big|\, {\sim}Ba \,\&\, {\sim}Ra\big) - P(H) = P(H)\left[\frac{P(H)}{P({\sim}Ba)P({\sim}Ba)} - 1\right].$$

Proponents of the standard solution assume that (1) $P(H) = P(Ra)$, and equivalently, *P*(*~Ra*|*H*) = *P*(*~Ra*). Likewise, they assume that (2) $P({\sim}Ba|H) = P({\sim}Ba)$. Let us apply these assumptions to the derived formula. Given (2), *P*(*~Ba*|*H*) and *P*(*~Ba*) cancel each other, which entails:

$$P\big(H \,\big|\, {\sim}Ba \,\&\, {\sim}Ra\big) - P(H) = P(H)\left[\frac{1}{P({\sim}Ba)} - 1\right].$$

Thus, so long as $0 < P({\sim}Ra|{\sim}Ba) < 1$ and $P(H) > 0$, it will follow that $P(H|{\sim}Ba \,\&\, {\sim}Ra) > P(H)$. In other words, it will follow that *~Ba & ~Ra* confirms *H*.

However, accepting (1) and (2) has an undesirable effect: *Ba & ~Ra* disconfirms *H* (Rinard 88-90, Maher 61, Vranas 551, and Horwich 58-9 have also made this claim). Consider the following chart representing our initial credences.

Initial Credences

	Ra	*~Ra*
Ba	*k*	*l*
~Ba	*m*	*n*

Suppose we have non-zero initial credences for each of the four possibilities represented in the chart, and the sum of the initial credences (*k, l, m,* and *n*) is 1. Now if we learn that all ravens are black, our credence for observing *Ra & ~Ba* will be 0. But this means our initial credence for it, *m,* must be somehow redistributed among the other three possibilities. If we accept both (1) and (2), then our credences for *Ba & Ra* and *~Ba & ~Ra* must each increase by *m,* otherwise $P(Ra) \neq P(Ra|H)$ and $P({\sim}Ba) \neq P({\sim}Ba|H)$. If that is all we do in updating our credences, the sum of our credences would be $1 + m$, which is unacceptable unless $m = 0$. So it follows that our updated credence for *~Ra & Ba* must decrease by *m.* The following table shows what our updated credences must be given (1) and (2).

Credences on *H,* (1), and (2)

	Ra	*~Ra*
Ba	$k + m$	$l - m$
~Ba	*0*	$n + m$

Two absurdities follow. First, a black non-raven disconfirms the hypothesis that all ravens are black. In fact, if our initial credence for observing a non-black raven is equal to our initial credence for observing a black non-raven, it follows that a black non-raven disproves *H* (because $l - m = 0$). Second, our initial credence for observing a non-black raven cannot be greater than our initial credence for observing a black non-raven, otherwise $l - m < 0$.

These consequences seem absurd, so we should reject (1) or (2). My own contention is that we should reject both. In the next section, I explore the logic behind Nicod's Condition and show that "all ravens are black" should not decrease *P*(*Ba & ~Ra*).

III. NICOD'S CONDITION AND INFINITIES

"All ravens are black" is true iff there are no non-black ravens in the world. Why does finding a black raven make it more likely that there are no non-black ravens? Why would the existence of a thing of a certain type be evidence for the non-existence of things of another type?[4]

Suppose there are exactly 10 objects in the world and we are testing the hypothesis that all ravens are black.[5] We can construe these objects as 10 opportunities to find a non-black raven that would falsify "all ravens are black." Each observed object that is not a non-black raven is tantamount to one fewer opportunity to falsify the hypothesis. So, each object we observe that is not a non-black raven confirms *H.*

More formally, we know that $P(\sim\exists x(Rx \,\&\, \sim Bx)) + P(\exists x(Rx \,\&\, \sim Bx)) = 1$.[6] Now suppose that $a_1, a_2, a_3, \ldots, a_{10}$ each uniquely denotes the 10 objects in the world. If there is a non-black raven, it must be one of these 10 objects.[7] So either one of the 10 objects is a non-black raven, or there are no non-black ravens. Thus, by the second axiom of probability theory, it follows that:

$$P((\sim Ba_1 \,\&\, Ra_1) \vee (\sim Ba_2 \,\&\, Ra_2) \vee (\sim Ba_3 \,\&\, Ra_3) \vee \ldots \vee (\sim Ba_{10} \,\&\, Ra_{10})) + P(\sim\exists x(\sim Bx \,\&\, Rx)) = 1.$$

For the sake of vividness, suppose that each object has an independent initial probability of 0.5 for being ($\sim Ba \,\&\, Ra$) and an independent initial probability of 0.5 for being $\sim(\sim Ba \,\&\, Ra)$. This generates the following results:

$$P((\sim Ba_1 \,\&\, Ra_1) \vee (\sim Ba_2 \,\&\, Ra_2) \vee (\sim Ba_3 \,\&\, Ra_3) \ldots (\sim Ba_{10} \,\&\, Ra_{10})) = 0.999;$$
$$P(\sim\exists x(\sim Bx \,\&\, Rx)) = 0.001.$$

Now suppose we learn that a_1 is not a non-black raven. This piece of evidence allows us to deduce that if there is a non-black raven, it is among the remaining 9 objects in the world. So we update our credence in the following way:

$$P((\sim Ba_2 \,\&\, Ra_2) \vee (\sim Ba_3 \,\&\, Ra_3) \ldots (\sim Ba_{10} \,\&\, Ra_{10})) + P(\sim\exists x(\sim Bx \,\&\, Rx)) = 1.$$

Updating based on the evidence generates the following results:

$$P((\sim Ba_2 \,\&\, Ra_2) \vee (\sim Ba_3 \,\&\, Ra_3) \ldots (\sim Ba_{10} \,\&\, Ra_{10})) = 0.998;$$
$$P(\sim\exists x(\sim Bx \,\&\, Rx)) = 0.002.$$

Thus, each object observed that is not a non-black raven increases the probability that there are no non-black ravens, i.e. that all ravens are black.[8] Consider the following analogy. Suppose a particular baseball player has a 0.5 probability of failing to hit any particular pitch. So upon walking up to the home plate, he has a 0.125 probability of striking out. If he fails to hit the first pitch, the probability that he will strike out increases to 0.25. In the same way, observing $\sim(\sim Ba_1 \,\&\, Ra_1)$ confirms "all ravens are black."

Now suppose that in addition to knowing that there are 10 objects in the world we also know that 3 of these objects (say, $a_2, a_3,$ and a_4) are ravens. So either a_2, or a_3, or a_4 is a non-black raven, or there are no non-black ravens. This implies:

$$P((\sim Ba_2 \,\&\, Ra_2) \vee (\sim Ba_3 \,\&\, Ra_3) \vee (\sim Ba_4 \,\&\, Ra_4)) + P(\sim\exists x(Rx \,\&\, \sim Bx)) = 1.$$

Again, assume that each raven has a 0.5 probability of being black. This yields the following results as our prior probabilities:

$$P((\sim Ba_2 \,\&\, Ra_2) \vee (\sim Ba_3 \,\&\, Ra_3) \vee (\sim Ba_4 \,\&\, Ra_4)) = 0.875;$$
$$P(\sim\exists x(\sim Bx \,\&\, Rx)) = 0.125.$$

In this case, $P(H)$ increases more significantly upon learning $Ra_2 \,\&\, Ba_2$ than upon learning $\sim(\sim Ba_1 \,\&\, Ra_1)$. Here are the posterior probabilities after updating on $Ra_2 \,\&\, Ba_2$:

$$P((\sim Ba_3 \,\&\, Ra_3) \vee (\sim Ba_4 \,\&\, Ra_4)) = 0.75;$$
$$P(\sim\exists x(\sim Bx \,\&\, Rx)) = 0.25.$$

Now suppose that instead of knowing that there are 3 ravens, what we know is that there are 3 non-black things. In this case, learning $\sim Ba_5 \,\&\, \sim Ra_5$ raises the probability of H just as much as learning $Ra_2 \,\&\, Ba_2$ did in the previous case. Under the specified conditions, the evidential import of $\sim Ba_5 \,\&\, \sim Ra_5$ is, logically speaking, no different from the evidential import of $Ra_2 \,\&\, Ba_2$.

To generalize, where n is the number of unobserved F's, the evidential import ($\Delta P(H)$)[9] of $Fx \,\&\, Gx$ can be calculated as the difference between the prior probability ($P(H)_i$) and posterior probability of H ($P(H)_f$):

$$\Delta P(H) = P(H)_f - P(H)_i$$
$$\Delta P(H) = P(Gx|Fx)^{n-1} - P(Fx)^n$$

So we see how BNC works: if there is a finite number F's, then each time we observe an instance of F that is a G, that is one fewer opportunity we have to disprove the hypothesis that all F's are G's. But what if there is an infinite number of unobserved F's?[10] Suppose this is the case. Let $a_1, a_2, a_3, \ldots$ denote objects in the world. By the second axiom of probability theory, we know:

$$P((Fa_1 \,\&\, \sim Ga_1) \vee (Fa_2 \,\&\, \sim Ga_2) \vee (Fa_3 \,\&\, \sim Ga_3) \vee \ldots) + P(\sim\exists x(Fx \,\&\, \sim Gx)) = 1$$

Now if each F has a non-zero probability of being $\sim G$, then the probability that at least one F is $\sim G$, given an infinity of F's, is 1.[11] The logical consequence is that given an infinity of F's, the initial value for $P(\sim\exists x(Fx \,\&\, \sim Gx))$ is 0.[12]

Imagine now that we learn $Fa_1 \,\&\, Ga_1$. This does nothing to change the value of $P(\sim\exists x(Fx \,\&\, \sim Gx))$. The value of $P((Fa_2 \,\&\, \sim Ga_2) \vee (Fa_3 \,\&\, \sim Ga_3) \vee (Fa_4 \,\&\, \sim Ga_4) \vee \ldots)$ is equal to the value of $P((Fa_1 \,\&\, \sim Ga_1) \vee (Fa_2 \,\&\, \sim Ga_2) \vee (Fa_3 \,\&$

$\sim Ga_3$) v …). This is because in both cases, we have an infinite number of unobserved *F's*.[13]

Recall our formula for Δ*P*(*H*). The claim here is where *n* approaches infinity, Δ*P*(*H*) approaches 0.

$$P(Gx|Fx)^{n-1} - P(Fx)^{n} = 0$$

So,

$$\Delta P(H) = 0$$

This is true because $0 \leq P(Gx|Fx) \leq 1$. Where $0 \leq P(Gx|Fx) < 1$, $P(Gx|Fx)^{n-1} - P(Gx|Fx)^{n} = 0 - 0$. Where $P(Gx|Fx) = 1$, $P(Gx|Fx)^{n-1} - P(Gx|Fx)^{n} = 1 - 1$. So, in the infinite cases, BNC does not work. At least, there is no *logical* reason to think that BNC is true in cases where we have an infinite number of unobserved *F's*. There may be other reasons to believe that BNC holds even with an infinite number of *F's* left to observe. For example, one may think that each *F* that is a *G* confirms the hypothesis that there is a law of nature that all *F's* are *G's*. However, setting aside such possibilities, there is no logical reason for BNC to be true given an infinite number of unobserved *F*'s. For this reason, we should revise BNC. I propose the following revision:

> *Finite Nicod's Condition* (FNC): For any randomly sampled object *x* (which we know nothing else about) and any properties *F* and *G*, the proposition that *x* has both *F* and *G* confirms the hypothesis that every *F* has *G* only if there is a finite number of unobserved objects that have *F*.

IV. Further Applications and Corollaries of Finite Nicod's Condition

In the previous section, I propose a revision of Nicod's Condition, FNC, to restrict confirmation to the domain of the finite. So the question, for Hempel's paradox, is whether one believes that there is infinitely many unobserved non-black objects (infinitely many non-G's). For example, an Epicurean may think that a white swan does not confirm "all non-black things are non-ravens." However, if one believes there is only finitely many unobserved non-black objects, FNC allows the observation of a white swan to confirm *H.*

The relationship between Nicod's Condition and infinities is especially instructive because infinities are limiting cases. The greater the number of non-black things, the weaker the evidential import of ~*Ba* & ~*Ra* will be ($\Delta P(H) = 0$). Consider a mathematical example to illustrate this. Goldbach's conjecture states that every even integer greater than 2 can be expressed as the sum of two primes. We have neither found a proof nor a counterexample for Goldbach's conjecture. Every even integer greater than 2 we've "observed" thus far can be expressed as the sum of two primes. But there are infinitely many even integers we have not observed. So, if FNC is true, Goldbach's conjecture is not confirmed via Nicod's Condition when we observe even integers that can be expressed as the sum of two primes.

The example of Goldbach's conjecture complicates matters for us. Initially, it seems unsurprising that where we have an infinite number of unobserved *F'*s and $P(Gx|Fx) < 1$, Fa_1 & Ga_1 would not confirm "all *F*'s are *G*'s" because the *P*(all *F*'s are *G*'s) = 0. But it is not irrational to have a non-zero initial credence for the truth of Goldbach's conjecture, due to the possible existence of an undiscovered proof. If so, then even in a case where the prior probability of "all *F*'s are *G*'s" is greater than 0, Fa_1 & Ga_1 would not confirm "all *F*'s are *G*'s" given an infinite number of unobserved *F'*s.

So we have a case where there is an infinite number of *F*'s (meaning, no confirmation via Nicod's Condition), the initial probability of Fa_1 & Ga_1 is less than 1, but somehow the prior probability of "all *F*'s are *G*'s" is greater than 0. This is puzzling, for if $P(Gx|Fx) < 1$, then *P*(all *F'*s are *G'*s) given an infinity of *F*'s should be 0. Perhaps the only plausible way out is to deny that $P(Gx|Fx)$ is static. Upon observing that some even integer 2k can be expressed as the sum of two primes, we become more confident that 2k + 2 can also be expressed as the sum of two primes. So, "2k conforms to Goldbach's conjecture" and "2k + 2 conforms to Goldbach's conjecture" are not independent. The interdependence between the two events can be accounted for by allowing the possibility that there is an undiscovered proof for Goldbach's conjecture. Since "2k conforms to Goldbach's conjecture" confirms the hypothesis, "there is an undiscovered proof for Goldbach's conjecture," it increases our credence for "2k+2 conforms to Goldbach's conjecture." Analogously, a black raven confirms "all ravens are black" and it also confirms "there is a law nature that all ravens have to be black." Thus, observing a black raven would make us more confident that the next raven we observe will be black. A non-black non-raven, however, confirms only "all ravens are black" but not necessarily the law of nature governing ravens. Plausibly, this is why we are initially inclined to reject PC and why we tend to think black ravens carry more evidential weight for *H* than non-black non-ravens.

NOTES

1. Technically, confirmation is a relation between sentences. So it is false to say that an object confirms a hypothesis. That being said, I will depart from official practice in much of this paper. It is often simpler to say that a black raven confirms *H* than to say that the sentence, "*a* is a black raven" confirms *H* and most readers, in my experience, prefer a pleasant reading experience rather than conformity to this particular practice.
2. Though labeled as "standard," the standard Bayesian solution has received many

criticisms from Bayesians (e.g. Rinard, and Vranas) who, in their critique, continue to consider it the standard Bayesian solution.

3. For example, our background information may include, "if there is a black raven, then there is a non-black raven." Under this condition, a black raven would falsify *H* instead of confirming it.

4. Popper and others in favor of falsificationism may reject NC, claiming that the observation of a black raven—a failure of falsification—does not generate confirmation at all. So the rest of us who are less pessimistic about confirmation could use a defense of NC, lest we be convinced by Popper.

5. Here, and for the rest of the paper, all the objects I discuss are stipulated as observable objects. Unobservable objects have no chance of being observed, so their existence or lack thereof should not affect our probabilities before or after observation.

6. This is just an instance of the axiom that $P(H) + P(\sim H) = 1$.

7. Formally, $\exists x(Rx \ \& \sim Bx) \equiv ((\sim Ba_1 \ \& \ Ra_1) \lor (\sim Ba_2 \ \& \ Ra_2) \lor (\sim Ba_3 \ \& \ Ra_3) \lor \ldots \lor (\sim Ba_{10} \ \& \ Ra_{10})$.

8. Even in a case where the object is replaced after being observed, so that we cannot be sure whether the second observed object is not identical to the first, the probability that all ravens are black increases.

9. I am borrowing scientific notation here where $\Delta\Phi$ means "change in Φ," which is the difference between initial Φ (Φ_i) and final Φ (Φ_f). (Double check with the author about the use of the variable phi here.

10. This is not a trivial distinction; one might claim that it is possible to observe an infinite number of things. For example, perhaps when one looks at a ruler, one observes all the points (an uncountable infinite, in this case) between one end of the ruler and the other.

11. In fact, so long as each *F* has some non-zero probability of being *~G,* the probability that one of them is *~G*, given an infinity of *F's*, is 1.

12. In a case where we replace the sampled object, we still get confirmation with repeated observations of *Fx* & *Gx.* This is because repeated observations get us closer to observing a larger portion of the class of *F*'s. For example, if there are 10 *F'*s and we observe one *F* at a time with replacement, we can be pretty confident we have observed all 10 after 100 observations. But this is not true with an infinite number of *F*'s. We are no closer to observing all the *F*'s if there is an infinite number of unobserved *F*'s.

13. Even if one holds that the initial value of $P(\sim\exists x(Fx \ \& \sim Gx))$ is not zero, but infinitesimal, the argument goes through so long as $P((Fa_2 \ \& \sim Ga_2) \lor (Fa_3 \ \& \sim Ga_3) \lor (Fa_4 \ \& \sim Ga_4) \lor \ldots (Fa_n \ \& \sim Ga_n)) = P((Fa_1 \ \& \sim Ga_1) \lor (Fa_2 \ \& \sim Ga_2) \lor (Fa_3 \ \& \sim Ga_3) \lor \ldots (Fa_n \ \& \sim Ga_n))$.

Works Cited

Alexander, H. G. "The Paradoxes of Confirmation." *British Journal for the Philosophy of Science* vol. 9 no. 35 (1958): 227–33.

Clarke, Roger. "'The Ravens Paradox' Is a Misnomer." *Synthese* vol. 175 no. 3 (2010): 427–40.

Good, Irving John. "The White Shoe Is a Red Herring." *British Journal for the Philosophy of Science* vol. 17 no. 4 (1967): 322.

Hempel, Carl Gustav. "Studies in the Logic of Confirmation." *Mind* vol. 54 no. 213 (1945):

1–26.
Horwich, Paul. *Probability and Evidence*. Cambridge: Cambridge University Press, 1982.
Mackie, John. "The Paradox of Confirmation." *British Journal for the Philosophy of Science* vol. 13 no. 52 (1963): 265–77.
Maher, Patrick. "Inductive Logic and the Ravens Paradox." *Philosophy of Science* vol. 66 no. 1 (1999): 50–70.
Nicod, Jean. *Foundation of Geometry and Induction*. London: Routledge, 1930.
Rinard, Susanna. "A New Bayesian Solution to the Paradox of the Ravens." *Philosophy of Science* vol. 81 no. 1 (2014): 81–100.
Vranas, Peter B. M. "Hempel's Raven Paradox: A Lacuna in the Standard Bayesian Solution." *British Journal for the Philosophy of Science* vol. 55 no. 3 (2004): 545–60.