

# The Paradox of Confirmation

Branden Fitelson

Department of Philosophy  
 Group in Logic and the Methodology of Science  
 &  
 Institute for Cognitive and Brain Sciences  
 University of California-Berkeley

branden@fitelson.org  
 http://fitelson.org/

- 1 Overview
- 2 Hempel, Goodman & Quine
  - Hempel's Original Formulation of the Paradox
  - The (Inconsistent!) Approach of Hempel and Goodman
  - Quine's Approach
- 3 Bayesian Approaches: Old & New
  - Some Background on Bayesian Confirmation
  - Traditional Bayesian Approaches (from Carnap to Vranas)
  - A Better Bayesian Approach (with Jim Hawthorne)
- 4 Hempel Meets Bayes
- 5 References

- **Nicod Condition (NC):** For any object  $x$  and any properties  $\phi$  and  $\psi$ , the proposition that  $x$  is both  $\phi$  and  $\psi$  confirms the proposition that every  $\phi$  is  $\psi$ . More formally:  
 $(\forall \phi)(\forall \psi)(\forall x)[\phi x \ \& \ \psi x \text{ confirms } (\forall y)(\phi y \supset \psi y)].$
- **Equivalence Condition (EC):** For any propositions  $H_1$ ,  $E$ , and  $H_2$ , if  $E$  confirms  $H_1$  and  $H_1$  is (*classically!* [14]) logically equivalent to  $H_2$ , then  $E$  confirms  $H_2$ . More formally:  
 If  $E$  confirms  $H_1$ , and  $H_1 \models H_2$ , then  $E$  confirms  $H_2$ .
- **Paradoxical Conclusion (PC):** The proposition that  $a$  is both nonblack and a nonraven confirms the proposition that every raven is black. More formally (arbitrary particular  $a$ ):  
 $\sim Ba \ \& \ \sim Ra \text{ confirms } (\forall x)(Rx \supset Bx).$

**Proof.** (1) By (NC),  $\sim Ba \ \& \ \sim Ra$  confirms  $(\forall x)(\sim Bx \supset \sim Rx)$ .  
 (2) By Logic,  $(\forall x)(\sim Bx \supset \sim Rx) \models (\forall x)(Rx \supset Bx)$ .  
 $\therefore$  (PC) By (1), (2), (EC),  $\sim Ba \ \& \ \sim Ra$  confirms  $(\forall x)(Rx \supset Bx)$ .

Hempel [7] & Goodman [6] *embraced* (NC), (EC) *and* (PC). They saw **no paradox**. They *explain away* the paradoxical *appearance*:

... in the seemingly paradoxical cases of confirmation, we are often not judging the relation of the given evidence  $E$  *alone* to the hypothesis  $H$  ... instead, we tacitly introduce a comparison of  $H$  with ...  $E$  in conjunction with ... additional ... information we ... have at our disposal.

Idea:  $E [\sim Ra \ \& \ \sim Ba]$  confirms  $H [(\forall x)(Rx \supset Bx)]$  *relative to*  $\top$ , but  $E$  doesn't confirm  $H$  *relative to some background*  $K \neq \top$ .

Question: *Which*  $K \neq \top$ ? Answer:  $K = \sim Ra$ . Idea: If you already know that  $\sim Ra$ , then observing  $a$ 's color won't tell you anything about the color of ravens. Distinguish the following two claims:

- (PC)  $\sim Ra \ \& \ \sim Ba$  confirms  $(\forall x)(Rx \supset Bx)$ , *relative to*  $\top$ .
- (PC\*)  $\sim Ra \ \& \ \sim Ba$  confirms  $(\forall x)(Rx \supset Bx)$ , *relative to*  $\sim Ra$ .

**Intuition (I).** (PC) is true, but (PC\*) is false. [*Why?*  $\sim Ra$  reduces the size of the set of possible *counterexamples* to  $(\forall x)(Rx \supset Bx)$  [11].]

Nice idea! Sadly, (I) is *inconsistent* with their confirmation *theory!*

Specifically, intuition (*I*) contradicts (evidential) *monotonicity*:

(M) *E* confirms *H* relative to  $\top \Rightarrow E$  confirms *H* relative to *any* *K* (caveat: provided *K* mentions no individuals not mentioned in *E* or *H*).

And, Hempelian confirmation *theory entails* (M), because:

- Hempel explicates '*E* confirms *H*' as '*E*  $\models$  *Z*', where *Z* is a sentence constructed from *E* and *H* in a certain way. [This explains the *caveat* regarding (M): the only individuals that can appear in *Z* are those which already appear in *E* and/or *H*.]
- There is no distinction (in classical deductive logic) between '*E* entails *Z*, relative to background theory *K*' and '*E* in conjunction with *K* entails *Z*, relative to  $\top$  (i.e., *simpliciter*)'.
- Classical entailment ( $\models$ ) is *monotonic*: If *E* (alone) entails *Z*, then so does *E* in conjunction with (i.e., relative to) *any* *K*.

But, if (M) is true, then (PC)  $\Rightarrow$  (PC\*). So, their *theory contradicts* their *intuitive* suggestion (*I*) that (PC) is true, but (PC\*) is false.

Quine [13] rejects (PC) but accepts (EC). So, he rejects (NC). He argues that  $\forall \phi$  and  $\forall \psi$  in (NC) must be *restricted in scope*:

(NC')  $(\forall \phi \in \mathbf{N})(\forall \psi \in \mathbf{N})(\forall x)[\phi x \ \& \ \psi x \text{ confirms } (\forall y)(\phi y \supset \psi y)]$

Quine calls properties  $\phi, \psi$  satisfying (NC') "*projectible*." He says that *natural kinds* are distinctively projectible in this sense.

Many (e.g., H & G) are inclined to follow Quine in restricting (NC) to "natural kinds" (e.g., "GRUE"). But, most (e.g., H & G) *reject* Quine's classification of  $\sim R$  and  $\sim B$  in particular as "unnatural".

Quine thinks *R* and *B* are "natural" ("projectible"). As a result, he thinks *Ra* & *Ba* confirms  $(\forall x)(Rx \supset Bx)$ . What Quine denies is step (1) of our Proof:  $\sim Ba$  &  $\sim Ra$  confirms  $(\forall x)(\sim Bx \supset \sim Rx)$ .

Some have accepted Quine's diagnosis (e.g., Kim [10] Quines psychological laws). I think Quine's diagnosis is off the mark.

But, Quine is right that: (i) (NC) is *false*; and, (ii) we need a *unified account* of *all* the confirmation paradoxes. However, (NC) is false *even for natural kinds*, so we'll need a *different* unified account.

Bayesianism assumes that *epistemically rational* degrees of belief (i.e., *credences* of rational agents) satisfy the probability calculus.

$\Pr(H | K)$  is the degree of belief that a rational agent with *background knowledge* *K* assigns to *H* (called the "prior" of *H*).

$\Pr(H | E \ \& \ K)$  is the degree of belief a rational agent with background knowledge *K* assigns to *H* *on the supposition that/upon learning with certainty that* *E* ("posterior" of *H*, on *E*).

Toy Example: Let *H* be the proposition that a card sampled from some deck is a ♠, and *E* be the proposition that the card is black.

Making standard assumptions about random sampling from decks (*K*),  $\Pr(H | K) = \frac{1}{4}$  and  $\Pr(H | E \ \& \ K) = \frac{1}{2}$ . So, relative to *K*, learning that *E* (or supposing that *E*) *raises the probability of* *H*.

**Def.** *E* confirms *H*, relative to *K* iff  $\Pr(H | E \ \& \ K) > \Pr(H | K)$ . I'll abbreviate this three-place confirmation relation as  $\mathfrak{C}(H, E | K)$ .

Important Note:  $\mathfrak{C}(H, E | K)$  is *not* monotonic in either *E* or *K*!

There are *many* logically equivalent ways of saying *E* confirms *H*, relative to *K*. Here are the three most common of these:

- $\mathfrak{C}(H, E | K)$  iff  $\Pr(H | E \ \& \ K) > \Pr(H | K)$ . [ $\frac{1}{2} > \frac{1}{4}$ ]
- $\mathfrak{C}(H, E | K)$  iff  $\Pr(E | H \ \& \ K) > \Pr(E | \sim H \ \& \ K)$ . [ $1 > \frac{1}{3}$ ]
- $\mathfrak{C}(H, E | K)$  iff  $\Pr(H | E \ \& \ K) > \Pr(H | \sim E \ \& \ K)$ . [ $\frac{1}{2} > 0$ ]

By taking differences, ratios, *etc.*, of the left/right sides of such inequalities, various confirmation *measures*  $\mathfrak{c}(H, E | K)$  emerge.

When  $\mathfrak{c}(H, E_1 | K) > \mathfrak{c}(H, E_2 | K)$ , we say that *E*<sub>1</sub> confirms *H* *more strongly than* *E*<sub>2</sub> does, relative to *K*, according to measure  $\mathfrak{c}$ .

Most Bayesian confirmation measures  $\mathfrak{c}$  satisfy the following:

(\*) If  $\Pr(H | E_1 \ \& \ K) > \Pr(H | E_2 \ \& \ K)$ , then  $\mathfrak{c}(H, E_1 | K) > \mathfrak{c}(H, E_2 | K)$ .

We'll make use of this fact about confirmation measures, when we discuss *comparative* Bayesian approaches to the paradox.

But, first, let's see how Bayesians *represent* the paradox ...

All Bayesian approaches begin by *precisifying* (NC) [and (PC)].

Since Bayesian confirmation is a 3-place relation [ $\mathcal{C}(H, E | K)$ ], we'll need a *quantifier* over the *implicit*  $K$ 's in (NC). 4 renditions:

$$(NC_w) \quad (\exists K)(\forall \phi)(\forall \psi)(\forall x)[\mathcal{C}((\forall y)(\phi y \supset \psi y), \phi x \& \psi x | K)]$$

$$(NC_\alpha) \quad (\forall \phi)(\forall \psi)(\forall x)[\mathcal{C}((\forall y)(\phi y \supset \psi y), \phi x \& \psi x | K_\alpha)]$$

$$(NC_\top) \quad (\forall \phi)(\forall \psi)(\forall x)[\mathcal{C}((\forall y)(\phi y \supset \psi y), \phi x \& \psi x | K_\top)]$$

$$(NC_s) \quad (\forall K)(\forall \phi)(\forall \psi)(\forall x)[\mathcal{C}((\forall y)(\phi y \supset \psi y), \phi x \& \psi x | K)]$$

$(NC_w)$  is *too weak* [ $K = (\forall \phi)(\forall \psi)(\forall x)[\phi x \& \psi x \supset (\forall y)(\phi y \supset \psi y)]$ ].

As we'll soon see,  $(NC_s)$  is *too strong* (it's demonstrably false).

Thus,  $(NC_\alpha)$  and  $(NC_\top)$  will be the salient renditions of (NC).

Next, we'll look carefully at two kinds of Bayesian approaches:

- **Qualitative.** Argue that *some* rendition of (NC) is false.
- **Comparative.** Argue  $c(H, Ra \& Ba | K_\alpha) > c(H, \sim Ba \& \sim Ra | K_\alpha)$ .

I.J. Good [4] gave the following counterexample to  $(NC_s)$ :

Let  $K$  be: Exactly one of the following two hypotheses is true: ( $H$ ) there are 100 black ravens, no nonblack ravens, and 1 million other things [viz.,  $(\forall x)(Rx \supset Bx)$ ], or ( $\sim H$ ) there are 1,000 black ravens, 1 white raven, and 1 million other things.

Let  $E$  be  $Ra \& Ba$  ( $a$  randomly sampled from universe). Then:

$$\Pr(E | H \& K) = \frac{100}{1000100} \ll \frac{1000}{1001001} = \Pr(E | \sim H \& K)$$

$\therefore K, R, B, a$  are such that *not*- $\mathcal{C}((\forall x)(Rx \supset Bx), Ra \& Ba | K)$ . And so Good's example is indeed a counterexample to  $(NC_s)$ .

Therefore,  $(NC_s)$  is false, and *even for "natural kinds"* (pace Quine). Similar examples will show that  $(PC_s)$  is also false.

Hempel [8] complains that Good's example is irrelevant to  $(NC_\top)$ .

Is this a fair complaint? [No!] Anyhow, Good responds to it ...

Here's Good's [5] attempt to meet Hempel's  $(NC_\top)$  Challenge:

Imagine an infinitely intelligent newborn baby having built-in neural circuits enabling him to deal with formal logic, English syntax, and subjective probability. He might argue, after defining a crow in detail, that it is initially extremely unlikely that there are any crows, and  $\therefore$  it is extremely likely that all crows are black ... [but] if there are crows, then there is a reasonable chance they are a variety of colours ... if he were to discover that a black crow exists he would consider  $[H]$  to be less probable than it was initially.

Even Good wasn't so confident about this "counterexample" to  $(NC_\top)$ . Maher [11] argues this is *not* a counterexample to  $(NC_\top)$ .

Maher [12] has recently provided a very compelling (Carnapian) counterexample to  $(NC_\top)$ , which is beyond our scope today.<sup>1</sup>

Most Bayesians don't *understand*  $(NC_\top)$ . Unlike Carnap [1], they have *no theory* of " $\Pr_\top$ " [or " $\mathcal{C}(H, E | \top)$ "]. So, they *abandon qualitative approaches* in favor of *comparative approaches*.

<sup>1</sup>Maher [12] shows that  $\Pr_\top(H | E) < \Pr_\top(H)$ , for some adequate Carnapian  $\Pr_\top$  functions. Hence,  $(NC_\top)$  is false for a Carnapian theory of " $\mathcal{C}(H, E | \top)$ ".

There have been *many* comparative Bayesian approaches to the paradox (see [15], [2], [9]). Here is a canonical characterization:

Assume that our *actual* background corpus  $K_\alpha$  is such that:

$$(4) \Pr(\sim Ba | K_\alpha) > \Pr(Ra | K_\alpha)$$

$$(5) \Pr(Ra | H \& K_\alpha) = \Pr(Ra | K_\alpha) \quad [ \therefore \Pr(\sim Ra | H \& K_\alpha) = \Pr(\sim Ra | K_\alpha) ]$$

$$(6) \Pr(\sim Ba | H \& K_\alpha) = \Pr(\sim Ba | K_\alpha) \quad [ \therefore \Pr(Ba | H \& K_\alpha) = \Pr(Ba | K_\alpha) ]$$

**Theorem.** Any  $\Pr$  satisfying (4), (5) and (6) will also be such that:

$$(7) \Pr(H | Ra \& Ba \& K_\alpha) > \Pr(H | \sim Ba \& \sim Ra \& K_\alpha).$$

$\therefore$  By  $(*)$ , the proposition that  $a$  is a black raven will (*actually*) confirm that all ravens are black *more strongly than* the proposition that  $a$  is a nonblack nonraven, *if* (4)–(6) hold for  $K_\alpha$ .

(4) is rather plausible (and it's uncontroversial in the literature).

(5) and (6) are problematic. I'll say more about them below. For now, it's worth noting that Hempel wouldn't have liked them.

Moreover, (4)–(6) are quite strong. They entail *far more than* (7).

Assumptions (4)–(6) *also* entail the following *qualitative* claims:

- (8)  $\Pr(H \mid Ra \ \& \ Ba \ \& \ K_\alpha) > \Pr(H \mid K_\alpha)$
- (9)  $\Pr(H \mid \sim Ba \ \& \ \sim Ra \ \& \ K_\alpha) > \Pr(H \mid K_\alpha)$
- (10)  $\Pr(H \mid Ba \ \& \ \sim Ra \ \& \ K_\alpha) < \Pr(H \mid K_\alpha)$

Hempel’s theory agrees with (8) and (9), since it also implies that  $Ra \ \& \ Ba$  and  $\sim Ba \ \& \ \sim Ra$  confirm  $H$ . But, Hempel’s theory also entails that  $Ba \ \& \ \sim Ra$  confirms  $H$ . So, (10) is *non-Hempel*ian.

These consequences of (4)–(6) are undesirable for two reasons:

- They preclude (4)–(6) from grounding a *purely comparative* approach [i.e., one that’s *neutral* on the truth of (8) and (9)].
- According to *many* commentators on the paradox (both Hempelians and non-Hempelians — see [15] for several references here), *even if* (8) and (9) are plausible, (10) *isn’t*.

It would be nice to have a *purely comparative* approach — one which does not *force* the Bayesian to accept *any* of (8)–(10)...

The problematic assumptions are the *independencies*: (5) & (6).

Vranas [15] gives various compelling objections to (5) & (6), and their standard rationales. He also suggests (6) is “for all practical purposes *necessary*” for the traditional Bayesian approaches.

This is misleading, since assumptions *far weaker than* (5) & (6) suffice [with (4)] for a *comparative* approach (see [3] for details).

Comparatively, (5) & (6) can be replaced by the *strictly weaker*:

$$(\ddagger) \Pr(H \mid Ra \ \& \ K_\alpha) \geq \Pr(H \mid \sim Ba \ \& \ K_\alpha)$$

( $\ddagger$ ) says:  $Ra$  confirms  $H$  *no less strongly* than  $\sim Ba$  does. This assumption is far more plausible than the independencies (5) & (6). None of the standard arguments against (5)/(6) apply to ( $\ddagger$ ).

Moreover, accepting (4) & ( $\ddagger$ ) is consistent with denying (or accepting) all three of the qualitative claims (8), (9), and/or (10).

Thus, a more plausible, *purely comparative* approach *is* possible.

Most contemporary Bayesians (except Maher [12]) *accept* (PC). In this sense, modern Bayesians are rather “Hempelian” at heart.

Hempel appeals to “tautological” vs “nontautological”  $\mathfrak{C}$  in his “explanation away”, which *contradicts* his (monotonic) *theory*.

I suggest that Hempel was actually rather Bayesian at heart, and that what he had in mind was something like this (for *some*  $K$ ):

$$(11) \ \mathfrak{c}(H, \sim Ba \ \& \ \sim Ra \mid K) > \mathfrak{c}(H, \sim Ba \ \& \ \sim Ra \mid \sim Ra \ \& \ K) = 0$$

Maher [11] develops a Carnapian account that is consistent with (11). Unfortunately, the traditional Bayesian accounts are *not*.

Our weaker assumptions (4) & ( $\ddagger$ ) are also consistent with (11).

So, I propose a Hempel-Bayes solution. First, we must *distinguish*  $\mathfrak{C}(H, \sim Ba \ \& \ \sim Ra \mid K_\alpha)$  [T?], and  $\mathfrak{C}(H, \sim Ba \ \& \ \sim Ra \mid \sim Ra \ \& \ K_\alpha)$  [F].

Second, *even if* it turns out that  $\mathfrak{C}(H, \sim Ba \ \& \ \sim Ra \mid K_\alpha)$ , it is not implausible that  $\mathfrak{c}(H, Ra \ \& \ Ba \mid K_\alpha) > \mathfrak{c}(H, \sim Ba \ \& \ \sim Ra \mid K_\alpha)$ .

- [1] R. Carnap, *Logical foundations of probability*, 1950.
- [2] J. Earman, *Bayes or bust?*, 1992.
- [3] B. Fitelson and J. Hawthorne, *How Bayesian confirmation theory handles the paradox of the ravens*, *Probability in Science* (Eells and Fetzer, eds.), 2005.
- [4] I.J. Good, *The white shoe is a red herring*, BJPS (1967).
- [5] ———, *The white shoe qua red herring is pink*, BJPS (1968).
- [6] N. Goodman, *Fact, fiction, and forecast*, 1954.
- [7] C. Hempel, *Studies in the logic of confirmation*, *Mind* (1945).
- [8] ———, *The white shoe: No red herring*, BJPS (1967).
- [9] C. Howson and P. Urbach, *Scientific reasoning: The Bayesian approach*, 1993.
- [10] J. Kim, *Multiple realization and the metaphysics of reduction*, PPR (1992).
- [11] P. Maher, *Inductive logic and the ravens paradox*, *Philosophy of Science* (1999).
- [12] ———, *Probability captures the logic of scientific confirmation*, *Contemporary Debates in the Philosophy of Science* (Christopher Hitchcock, ed.), 2004.
- [13] W.V.O. Quine, *Natural kinds, Ontological Relativity and other Essays*, 1969.
- [14] R. Sylvan and R. Nola, *Confirmation without paradoxes*, *Advances in Scientific Philosophy* (G. Schurz and G. Dorn, eds.), 1991.
- [15] P. Vranas, *Hempel’s raven paradox: a lacuna in the standard Bayesian solution*, *British Journal for the Philosophy of Science* (2004).