

HEMPEL'S RAVEN PARADOX: A LACUNA IN THE STANDARD BAYESIAN SOLUTION

Peter B. M. Vranas
vranas@iastate.edu
Iowa State University
7 July 2003

Abstract. According to Hempel's paradox, evidence (E) that an object is a nonblack nonraven confirms the hypothesis (H) that every raven is black. According to the standard Bayesian solution, E does confirm H but only to a minute degree. This solution relies on the almost never explicitly defended assumption that the probability of H should not be affected by evidence that an object is nonblack. I argue that this assumption is implausible, and I propose a way out for Bayesians.

1. Introduction.¹

"What do you like best about philosophy?" my grandmother asks. "Every philosopher is brilliant!" I unhesitatingly reply. "That's hardly believable" my grandmother responds. To convince her I perform an experiment. I choose randomly a person, Rex, from the telephone directory. Predictably enough, Rex is neither brilliant nor a philosopher. "You see," I crow, "this confirms my claim that every philosopher is brilliant!" My grandmother is unimpressed.

I repeat the experiment: I choose randomly another person, Kurt. Surprisingly, Kurt is both a philosopher and brilliant. Now my grandmother is really impressed. "Look, grandma," I patiently explain, "you are being irrational. The fact that Kurt is both a philosopher and brilliant confirms the hypothesis that every philosopher is brilliant—or so you accept. But then the fact that Rex is both nonbrilliant and a nonphilosopher confirms the hypothesis that everything nonbrilliant is a nonphilosopher, and thus also confirms the *logically equivalent* hypothesis that every philosopher is brilliant—or so you should accept."

"Young man, you are being impertinent" my grandmother retorts. "You just rehearsed Hempel's paradox; I do remember my philosophy classes. But it's called a 'paradox' precisely because its conclusion is absurd."

"Is it, though?" I shamefacedly reply. "Not according to Bayesians. They argue that the fact that Rex is neither brilliant nor a philosopher does confirm the hypothesis that every philosopher is brilliant—although only to a minute degree, this is why we have the illusion that it doesn't confirm it at all."

"Young man, you are being superficial" my grandmother responds. "Go through the Bayesian argument. It relies on the assumption that the fact that Rex is not brilliant should not affect the

I am very grateful to Paul Bartha, Michael Bishop, Branden Fitelson, Alan Hájek, and Patrick Maher for helpful comments. Thanks also to Jack Davidson, Kevin de Laplante, Carla Fehr, Malcolm Forster, Heimir Geirsson, Aviv Hoffmann, William Robinson, Brian Skyrms, Tony Smith, and Richard Swinburne for help; and to James Joyce, Peter Railton, and Lawrence Sklar for comments on a distant ancestor of this paper. A version of this paper was presented at the Eighteenth Biennial Meeting of the Philosophy of Science Association (Milwaukee, November 2002).

¹ This section presents the issues in a simplified and slightly imprecise way. Rigor is introduced in later sections.

probability of the hypothesis that every philosopher is brilliant. What on earth justifies this assumption?”

I’m dumbfounded. What, indeed? Apparently nothing that Bayesians themselves have said: the assumption in question is almost never explicitly defended in the literature. After some serious thought, I conclude that my grandmother is right: the assumption is implausible. The standard Bayesian solution to Hempel’s paradox has a lacuna.

In §2 I formulate more carefully Hempel’s paradox, the standard Bayesian solution, and the disputed assumption. In §3 and §4 I examine respectively arguments for and against the assumption. In §5 I propose a way out for Bayesians. I conclude in §6.

2. Hempel’s paradox, the standard Bayesian solution, and the disputed assumption.

Hempel’s paradox can be formulated as the argument from NC and EC to PC:²

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 proposition that every F has G .³

Equivalence Condition (EC): For any propositions P , Q , and Q' , if P confirms Q and Q is logically equivalent to Q' , then P confirms Q' .⁴

Paradoxical Conclusion (PC): The proposition (E) that a is both nonblack and a nonraven confirms the proposition (H) that every raven is black.⁵

The argument is valid: the proposition that a is both nonblack and a nonraven confirms (by NC) the proposition that everything nonblack is a nonraven, and thus confirms (by EC) the logically equivalent proposition that every raven is black.⁶

² The exact origin of Hempel’s paradox is shrouded in mystery. Although Hempel apparently did not formulate the paradox in print until 1943 (128), Hosiasson-Lindenbaum formulated it as early as 1940 (136); she attributed it to Hempel but gave no reference. (Hempel (1945: 21 n. 2) referred to “discussions” with her.) The paradox was “foreshadowed” (Jeffrey 1995: 3) but by no means formulated by Hempel in 1937 (222).

³ NC is part of Nicod’s (1924: 23, 1930: 219) criterion of confirmation, which also includes (inter alia) a claim that entails the negation of PC: the proposition that x has neither F nor G does not confirm the proposition that every F has G . Some authors include that claim (or directly the negation of PC) in the premises of Hempel’s paradox; they formulate thus the paradox as an apparently inconsistent set of plausible premises rather than (as I do) as an apparently sound argument with an apparently unacceptable conclusion.

⁴ Some authors reject or modify EC. On the relevant debate see: Black 1966: 186; Fisch 1984; Foster 1971: 107-10; Giere 1970: 359, 361; Goodman 1983: 70-1; Grandy 1967: 22-3; Hempel 1945: 12-3; Lipton 1991: 102-3, 105; Morgenbesser 1962; Rescher 2001: 225-6; Rody 1978: 298-300; Scheffler 1963: 286-91; Scheffler & Goodman 1972: 80-2; Schoenberg 1964: 202-8; Schwartz 1972: 247-8; Skyrms 1966: 242-3; Smokler 1967; Swinburne 1971: 320-1; Tempelmeier 1972: 10-1; Tuske 1998: 391, 399-400; Vincent 1975: 3-18; Young 1975: 56-60.

⁵ People sometimes speak as if it were a itself (the *object*) that supposedly confirms H . As Hempel notes, however, “[a] particular bird may be a crow and black, but may also have an albino crow for a sister; in virtue of these properties, it would both confirm and disconfirm the hypothesis ‘All crows are black’. This consideration suggests that an object can be said to confirm or to disconfirm a hypothesis only *under a particular description*” (1967: 239; cf. Stove 1966: 446). People also sometimes say that, because (e.g.) a green leaf is a nonblack nonraven, H is confirmed (according to a variant of Hempel’s paradox) by the proposition that a is a green leaf. Apparently they presuppose that, if P confirms Q and P' entails P , then P' confirms Q . But then H would be both confirmed and disconfirmed by the proposition that a is a green leaf and b is a nonblack raven (Stillwell 1985; Stove 1966).

⁶ In *nomological* versions of Hempel’s paradox, in which the sentences “every raven is black” and “everything nonblack is a nonraven” are used to express putative laws (rather than, e.g., accidental generalizations), one might deny that these sentences express logically equivalent propositions. On the relevant debate see: Black 1966: 186-7; Cohen 1987: 158-60; Gaifman 1979: 115; Harré 1970: 120-1; Hempel 1945: 15-8; Huggett 1960; Nerlich 1964; Skyrms 1966:

The standard Bayesian solution to the paradox tries to vindicate PC. Bayesians argue that E does confirm H —but only to a minute degree, given that there are overwhelmingly more nonblack objects than there are ravens. Bayesians also claim that PC *looks* unacceptable (i.e., we have the impression that E does not confirm H at all) because we implicitly realize that the degree to which E confirms H is for all practical purposes negligible (and is much smaller than the degree to which the proposition that a is both black and a raven confirms H).⁷

More formally, let R , \bar{R} , B , and \bar{B} be respectively the properties of being a raven, a nonraven, black, and nonblack; then $H \equiv \forall x(Rx \rightarrow Bx)$ and $E \equiv \bar{B}a\bar{R}a$. Let the degree to which E confirms H be $P(H|E) - P(H)$,⁸ P being a subjective probability measure. It can be shown that:

$$(1) \quad P(H|E) - P(H) = P(H) \left[\frac{P(\bar{B}a|H) / P(\bar{B}a)}{P(\bar{R}a|\bar{B}a)} - 1 \right].^9$$

If it is part of the background knowledge that a is drawn randomly from a population containing overwhelmingly more nonblack objects than ravens, then $P(Ra) / P(\bar{B}a)$ is minute¹⁰—and thus so is $P(Ra|\bar{B}a)$, which is by definition $P(Ra\bar{B}a) / P(\bar{B}a)$ and thus does not exceed $P(Ra) / P(\bar{B}a)$. Say, then, that $P(Ra|\bar{B}a)$ is equal to ε^2 (with ε minute), and substitute in (1) $1 - \varepsilon^2$ (namely $1 - P(Ra|\bar{B}a)$) for $P(\bar{R}a|\bar{B}a)$.

Bayesians standardly assume that $P(Ba|H)$ and $P(Ba)$ should be equal; this is the *disputed assumption*. Equivalently, $P(\bar{B}a|H)$ and $P(\bar{B}a)$ (and thus also $P(H|\bar{B}a)$ and $P(H)$) should be equal. Then (1) gives that $P(H|E) - P(H)$ should be $P(H)[(1 - \varepsilon^2)^{-1} - 1]$, namely $P(H)\varepsilon^2/(1 - \varepsilon^2)$; this is both positive (assuming $P(Ra|\bar{B}a) > 0$) and minute, just as Bayesians claim. It follows that, in the presence of the other assumptions made so far, the disputed assumption is *sufficient* for the

235-6, 242-3; Stove 1966: 452; Swinburne 1971: 318-20; von Wright 1966: 217-8; Wilson 1964a, 1964b: 400-1; Young 1975: 55-6.

⁷ I refer to the *standard* Bayesian solution because “there is no such thing as *the* Bayesian solution. There are many different ‘solutions’ that Bayesians have put forward using Bayesian techniques” (Chihara 1981: 448). See, e.g.: Alexander 1958: 230-3; Chihara 1981: 440-8; Earman 1992: 69-73; Eells 1982: 60-1; Gaifman 1979; Gibson 1969; Good 1960, 1961; Hesse 1974: 155-62; Hooker & Stove 1968; Horwich 1982: 54-63; Hosiasson-Lindenbaum 1940: 136-41; Howson & Urbach 1993: 126-130; Jardine 1965; Mackie 1963: 266-9; Nerlich 1964; Suppes 1966; Swinburne 1971: 322-7, 1973: chap. 10; Wilson 1964b: 396-9; Woodward 1985: 409-16. Cf. Hintikka 1969; Humburg 1986; Maher 1999.

⁸ This is not the only measure of degree of confirmation that Bayesians have used (Fitelson 1999, 2001), but it can be shown that the sufficiency and necessity results of this section (see below in the text) also hold for, e.g., the log-ratio measure $\log[P(H|E)/P(H)]$, the log-likelihood-ratio measure $\log[P(E|H)/P(E|\bar{H})]$ (assuming for the sufficiency result that $P(\bar{H})$ is non-minute), and the normalized difference measure $P(H|E) - P(H|\bar{E}) = [P(H|E) - P(H)] / P(\bar{E})$ defended by Christensen (1999; cf. Eells & Fitelson 2000; Joyce 1999: 205-6) (assuming for the sufficiency result that $P(Ba)$ is non-minute and for the necessity result that $P(H)$ is non-minute).

⁹ *Proof.* By Bayes’ theorem, $P(H|E) = P(E|H)P(H) / P(E)$. Now $P(E) = P(\bar{B}a\bar{R}a) = P(\bar{R}a|\bar{B}a)P(\bar{B}a)$, and $P(E|H) = P(\bar{B}a\bar{R}a|H) = P(\bar{R}a|\bar{B}aH)P(\bar{B}a|H) = P(\bar{B}a|H)$ because $P(\bar{R}a|\bar{B}aH) = 1$. So $P(H|E) = P(\bar{B}a|H)P(H) / [P(\bar{R}a|\bar{B}a)P(\bar{B}a)]$, and (1) quickly follows. \square (This proof assumes $P(E)P(H)P(\bar{B}aH) > 0$.)

¹⁰ For the sake of simplicity I don’t consider—except briefly in footnote 18—cases in which a is drawn from a population of (e.g.) only nonravens or only black objects (cf. Black 1966: 184; Horwich 1982: 58; Jardine 1965: 361-3; Royall 1997: 177-9; Woodward 1985: 411). For random sampling to be possible, assume that the population is finite. The inference to minuteness requires that P satisfy something like David Lewis’s (1980, 1994) Principal Principle.

Bayesian claim that E confirms H to a minute degree. To my knowledge, however, it has escaped notice that the assumption is also for all practical purposes *necessary* for the Bayesian claim. (1) entails that $P(H|E) - P(H)$ is positive exactly if $P(\bar{B}a|H) / P(\bar{B}a) > 1 - \varepsilon^2$, and is minute—say less than some minute number δ^2 —exactly if $P(\bar{B}a|H) / P(\bar{B}a) < (1 - \varepsilon^2)[1 + (\delta^2/P(H))]$. Assuming that $P(H)$ is non-minute, it follows that $P(H|E) - P(H)$ is both positive and minute exactly if $P(\bar{B}a|H) / P(\bar{B}a)$ is very close to 1; and this holds only if $P(Ba|H) \cong P(Ba)$ (i.e., only if $|P(Ba|H) - P(Ba)|$ is minute). I conclude that the disputed assumption is for all practical purposes necessary for the standard Bayesian solution to work.¹¹

3. Attempts to defend the disputed assumption.¹²

Despite being essential for the standard Bayesian solution, the disputed assumption is almost never explicitly defended in the literature.¹³ The only argument that I have encountered in support of the assumption was adduced by Woodward: “in the absence of some special reason for supposing otherwise, it seems reasonable that my estimate of the number of masses in the universe should not go down (or up) when I learn that they all obey the inverse square law” (1985: 415).¹⁴ More formally, let P and P_+ be respectively my subjective probability measures right before and right after I learn that H is true. Woodward’s argument might be formalized as follows: (1) $P_+(Ba)$ and $P(Ba)$ should be equal; (2) $P_+(Ba)$ and $P(Ba|H)$ should be equal; thus (3) $P(Ba|H)$ and $P(Ba)$ should be equal. How convincing, however, is premise (1)? One person’s modus ponens is another’s modus tollens: granting (2), someone who denies (3) will find (1) implausible. In other words, if my prior probability measure P is (rationally) such that $P(Ba|H)$ and $P(Ba)$ differ, then (given (2)) $P_+(Ba)$ and $P(Ba)$ should differ—my estimate of the percentage of black objects *should* go up or down when I learn that they include every raven. Of course for all I have said it may still be the case that it’s epistemically irrational to have a prior which violates (3); my present point is only that Woodward’s argument doesn’t *show* this to be the case.

Another attempt—this one not in the literature—to defend the disputed assumption appeals to the following *Principle of Conditional Indifference* (PCI): if none of H_1, H_2 gives more reason than the

¹¹ In addition to claiming that (i) E confirms H to a minute degree, Bayesians standardly claim that (ii) $BaRa$ confirms H to a non-minute degree, or—equivalently given (i)—that (iii) the degree to which $BaRa$ confirms H is much larger than (i.e., exceeds by a non-minute amount) the degree to which E confirms H . I am not claiming that the disputed assumption is necessary (or sufficient) for (ii) or (iii), but my result that the assumption is for all practical purposes necessary for (i) entails my main claim that the assumption is for all practical purposes necessary for the standard Bayesian solution as a whole.

¹² One might argue that the disputed assumption is trivially true: the percentage of black objects in the population is *fixed* and thus cannot depend on H . But one might equally well (or rather equally badly) argue that the percentage of ravens that are black is fixed and thus cannot depend on H ! These percentages are fixed but *unknown*; as far as our state of knowledge is concerned, the situation is as if the population had been drawn from a set of possible populations, and the percentage of black objects does vary across this set.

¹³ Gaifman claims that, “not having ... a background theory, given only a natural classification into A ’s and non- A ’s and into B ’s and non- B ’s, there is no way in which knowing A ’s frequency can be relevant to estimating B ’s frequency within A ” (1979: 126). Howson and Urbach comment that the assumption “seems plausible to us, at any rate as a good approximation” (1993: 127). That’s about all the “defense” of the assumption that I have found in the literature.

¹⁴ Strictly speaking, Woodward—similarly for Howson and Urbach (footnote 13)—is discussing the assumption that $P(Ra|H)$ and $P(Ra)$ should be equal (an assumption that Bayesians standardly make to support claim (ii) of footnote 11), but his discussion applies to the disputed assumption as well.

other to believe E , then $P(E|H_1)$ and $P(E|H_2)$ should be equal. Whether or not every raven is black seems irrelevant to whether the (randomly selected) object a is black, so PCI gives that $P(Ba|H)$ and $P(Ba|\bar{H})$ should be equal; this is equivalent to the disputed assumption. PCI, however, bears a suspicious resemblance to the following *Principle of Indifference* (PI): if E gives no more reason to believe one of H_1, H_2 rather than the other, then $P(H_1|E)$ and $P(H_2|E)$ should be equal. PI is well known to “yield inconsistencies with alarming ease” (Howson & Urbach 1993: 59), but one might argue that PCI does not do so. PI yields an inconsistency if it is applied to two partitions of logical space, $\{H_1, \dots, H_n\}$ and $\{H'_1, \dots, H'_m\}$, which have different numbers of members ($n \neq m$) but share a member—say $H_1 = H'_1$. Indeed, PI in such a case entails that $P(H_1|E)$ should be equal to both $1/n$ and $1/m$. PCI, by contrast, in such a case entails that $P(E|H_1)$ should be equal to $P(E)$,¹⁵ a value independent of the partition. Nevertheless, in the Appendix I argue that PCI does lead to inconsistency in the context of the disputed assumption. So this attempt to defend the assumption by using indifference fails.

There is a general reason why the disputed assumption is hard to defend. Suppose one somehow refutes the claim that my estimate of the percentage of black objects should go up or down when I learn that they include every raven. The disputed assumption does not follow: it does not follow that my estimate should remain the same. What follows instead is that my estimate *may* remain the same. Indeed: denying that my estimate *should* go up or down is compatible with affirming that it *may* go up or down and thus does not entail that it should remain the same. So even if there is no reason why $P(Ba|H)$ and $P(Ba)$ should differ, maybe there is no reason why they should be equal either: maybe they may differ and they may also be equal. But why understand the disputed assumption as the claim that $P(Ba|H)$ and $P(Ba)$ *should* be—rather than *may* be or *are*—equal? Because the assumption is used to defend the claim that E confirms H , and this is most reasonably understood as the claim that $P(H|E)$ *should* exceed $P(H)$; as Horwich (1982: 52) puts it, “reason requires” that $P(H|E)$ exceed $P(H)$.¹⁶ I conclude that the disputed assumption makes a relatively strong claim (“should” rather than “may”) and is for that reason hard to defend. In any case, it seems that no adequate defense of the assumption exists.

4. Attempts to refute the disputed assumption.¹⁷

Opponents of the disputed assumption have been more diligent than proponents: I have found in the literature three attempts to refute the assumption. I will argue, however, that all three attempts fail.

¹⁵ *Proof.* $P(E) = P(E|H_1)P(H_1) + \dots + P(E|H_n)P(H_n) = P(E|H_1)[P(H_1) + \dots + P(H_n)] = P(E|H_1)$. \square

¹⁶ As Woodward (1985: 412-3) notes, this understanding of confirmation has the consequence that when E neither confirms nor disconfirms H it doesn't follow that E is confirmationally irrelevant to H (i.e., that $P(H|E)$ and $P(H)$ should be equal). But I take it that's unavoidable: replacing “should exceed” with “exceeds” makes confirmation “too subjective” (Horwich 1982: 51), and replacing “should exceed” with “may exceed” has the consequence that E can *both* confirm and disconfirm H .

¹⁷ One might argue that Bayesians should take the disputed assumption to be trivially false: all that epistemic rationality requires according to Bayesians is conformity of degrees of belief to the probability axioms (and modification of degrees of belief by Bayesian conditionalization), so epistemic rationality according to Bayesians cannot require that $P(Ba|H)$ and $P(Ba)$ be equal. In reply I contest the premise of this argument: Bayesians can accept that epistemic rationality also requires conformity to, e.g., Lewis's (1980, 1994) Principal Principle (cf. footnote 10) or van Fraassen's (1984, 1995) Reflection Principle. Or even conformity to some indifference principle which, like PCI (see §3), entails that $P(Ba|H)$ and $P(Ba)$ should be equal; I don't see how to exclude this possibility a priori.

An indirect attempt to refute the disputed assumption was made by Horwich (1982: 58-9). He argued that the assumption has the counterintuitive consequence that H is slightly *disconfirmed* by the proposition that a is both black and a *nonraven*.¹⁸ This consequence was called “awkward” by Swinburne (1971: 324, 1973: 158), “a minor embarrassment” by Horwich (1982: 58), “unintuitive” by Maher (1999: 61), “a paradox” by Hooker and Stove (1968: 307), “unacceptable” by Pennock (1991: 66 n. 21), and “a debilitating weakness” by Rody (1978: 289). I don’t see, however, why Bayesians should be embarrassed. To the charge that on their account $\overline{Ba}\overline{Ra}$ confirms H , Bayesians reply that the degree of confirmation is minute. Similarly, as Horwich himself points out, to the charge that on their account $Ba\overline{Ra}$ disconfirms H , Bayesians can reply that the degree of disconfirmation is minute.¹⁹ If one accepts the first reply, then why reject the second?²⁰

Another attempt to refute the disputed assumption was made by Swinburne: “But now h , ‘all ϕ ’s are ψ ’ is added to our evidence. This is going to lead us to suppose that there are more ψ ’s and less ϕ ’s than, without it, we had supposed” (1971: 325). This argument is analogous to Woodward’s (see §3) and might be similarly formalized: (1’) $P_+(Ba)$ should exceed $P(Ba)$; (2) $P_+(Ba)$ and $P(Ba|H)$ should be equal; thus (3’) $P(Ba|H)$ should exceed $P(Ba)$. I reply again that one person’s modus ponens is another’s modus tollens: if my prior probability measure P is (rationally) such that $P(Ba|H)$ and $P(Ba)$ are equal, then (given (2)) $P_+(Ba)$ and $P(Ba)$ should be equal. Again, for all I have said it may still be the case that it’s epistemically irrational to have a prior which violates (3’), but my present point is only that Swinburne’s argument doesn’t *show* this to be the case.²¹

¹⁸ Horwich (1982: 58-9), like Kruse (2000), also claims that the disputed assumption has the consequence that Bayesians are unable to distinguish cases in which a is drawn from different populations (cf. footnote 10). This is not so, however. Consider, for example, the case in which it is part of the background knowledge that a is drawn randomly from the *nonravens* in the population we have been considering so far. (H is still the proposition that every raven in the original population—not, of course, in the subpopulation of nonravens—is black.) Let P' be a corresponding subjective probability measure and assume that $P'(Ba|H)$ and $P'(Ba)$ should be equal. Then $P'(H|\overline{Ba}) = P'(\overline{Ba}|H)P'(H)/P'(\overline{Ba}) = P'(H)$, so we do get the intuitively correct result that \overline{Ba} is confirmationally irrelevant to H (because in this experiment H is at no risk of being falsified).

¹⁹ *Proof.* $P(H|Ba\overline{Ra}) = P(Ba\overline{Ra}|H)P(H) / P(Ba\overline{Ra}) = [P(Ba|H) - P(BaRa|H)]P(H) / P(Ba\overline{Ra}) = [P(Ba|H) - P(Ba|RaH)P(Ra|H)]P(H) / P(Ba\overline{Ra}) = [P(Ba|H) - P(Ra|H)]P(H) / P(Ba\overline{Ra}) = [P(Ba) - P(Ra)]P(H) / P(Ba\overline{Ra})$ if $P(Ba|H) = P(Ba)$ and (see footnote 14) $P(Ra|H) = P(Ra)$. So $[P(H|Ba\overline{Ra}) - P(H)]P(Ba\overline{Ra}) / P(H) = P(Ba) - P(Ra) - P(Ba\overline{Ra}) = P(BaRa) - P(Ra) = -P(\overline{Ba}Ra)$. Thus $P(H|Ba\overline{Ra}) - P(H) = -P(H)P(\overline{Ba}Ra) / P(Ba\overline{Ra}) = -P(H)P(Ra|\overline{Ba})P(\overline{Ba}) / [P(Ba) - P(BaRa)] = -P(H)P(Ra|\overline{Ba})P(\overline{Ba}) / [P(Ba)(1 - P(Ra|Ba))]$, which is minute if the background knowledge ensures that $P(Ra|\overline{Ba})$ and $P(Ra|Ba)$ are minute but $P(Ba) / P(\overline{Ba})$ is non-minute (there are overwhelmingly more nonblack objects and overwhelmingly more black objects than ravens, but at most many—not overwhelmingly—more nonblack than black objects). □ (Similar proofs can be given for the other three confirmation measures mentioned in footnote 8.)

²⁰ Of course some people reject the first reply. On the relevant debate see: Aronson 1989: 232; Black 1966: 196; Chihara 1981: 442; Cohen 1987: 154; Hempel 1965: 48; Hintikka 1969: 27; Poundstone 1988: 39.

²¹ Swinburne tries to bolster his argument as follows: “If all future ϕ ’s are ψ , then in so far as ϕ ’s are similar to $\sim\phi$ ’s, future $\sim\phi$ ’s are more likely to be ψ than we would otherwise have supposed” (1971: 325; cf. Maher 2002: 7). But why suppose that ϕ ’s are similar to $\sim\phi$ ’s?

A third attempt to refute the disputed assumption was made by Maher: “There are just two ways \overline{Ba} can be true,” namely $\overline{Ba}Ra$ and $\overline{Ba}\overline{Ra}$, and H rules out $\overline{Ba}Ra$; “since [H] does not rule out any of the ways in which a can be black, it is plausible that” $P(\overline{Ba})$ should exceed $P(\overline{Ba}|H)$ (1999: 60). Maher’s conclusion is equivalent to the claim that $P(\overline{Ba}|\overline{H})$ should exceed $P(\overline{Ba}|H)$, which is in turn equivalent to the claim that $P(\overline{Ba}Ra|\overline{H}) + P(\overline{Ba}\overline{Ra}|\overline{H})$ should exceed $P(\overline{Ba}Ra|H) + P(\overline{Ba}\overline{Ra}|H)$. Maher points out that $P(\overline{Ba}Ra|H)$ should be zero. Does then Maher’s argument amount to no more than the (clearly unwarranted) assertion that, because the former of the last two sums has *two* potentially positive terms but the latter has just *one*, the former should exceed the latter?²² Maher might note in response that the only potentially positive term in the latter sum, namely $P(\overline{Ba}\overline{Ra}|H)$, “corresponds” to one of the terms in the former sum, namely $P(\overline{Ba}\overline{Ra}|\overline{H})$. But how would this correspondence help Maher’s argument, unless he assumed that the corresponding terms should be equal? And clearly he would not be entitled to make this assumption (I am not saying he does make it), which is equivalent to the claim that $P(H|\overline{Ba}\overline{Ra})$ and $P(H)$ should be equal: this would beg the question against Bayesians, who claim that $\overline{Ba}\overline{Ra}$ does confirm H . I conclude that Maher’s argument fails.²³

Where does this leave us? I think that Swinburne’s and Maher’s arguments were bound to fail because they tried to show too much, namely that $P(Ba|H)$ and $P(Ba)$ *should* differ.²⁴ To refute the disputed assumption it’s enough to show instead that these two quantities *may* differ, and to show this one could try to rebut every putative reason why they should be equal. In §3 I rebutted the only two such reasons of which I am aware. Although the disputed assumption is thereby made implausible, it is clearly not decisively refuted. Still, until further reasons are adduced, it seems sensible for those Bayesians who rely on the assumption to look for a way to stop relying on it.

5. A way out for Bayesians.

It is important to distinguish two questions that a full solution to Hempel’s paradox must address:

Prescriptive question: Should $P(H|\overline{Ba}\overline{Ra})$ exceed $P(H)$ (i.e., does $\overline{Ba}\overline{Ra}$ confirm H)?

Explanatory question: Why do people believe that $P(H|\overline{Ba}\overline{Ra})$ and $P(H)$ should be equal (i.e., that $\overline{Ba}\overline{Ra}$ is confirmationally irrelevant to H)?

²² Strictly speaking, since Maher compares the implications of H for \overline{Ba} and for Ba , one might have expected him to conclude (not that $P(\overline{Ba}|\overline{H})$ should exceed $P(\overline{Ba}|H)$, but rather) that $P(Ba|H)$ should exceed $P(\overline{Ba}|H)$; equivalently, that $P(BaRa|H) + P(\overline{Ba}\overline{Ra}|H)$ should exceed $P(\overline{Ba}Ra|H) + P(\overline{Ba}\overline{Ra}|H)$, a claim to which the (dubious) reasoning of *two* versus *one* potentially positive terms also applies. But each of the last two claims is equivalent to the claim that $P(Ba|H)$ should exceed 1/2, so this reconstruction of Maher’s argument might be too uncharitable.

²³ Maher also argues (1999: 60) that the disputed assumption is incompatible with imposing five conditions on P (including Carnap’s “ λ -condition”). Maher, however, admits he “cannot prove” (1999: 53) that these conditions hold in the context of Hempel’s paradox.

²⁴ Forster (1994: 362-3), following Good (1967: 322), argues that $P(Ra|H)$ and $P(Ra)$ should differ in the presence of appropriate background knowledge; but it does not follow that they should differ in the *absence* of such knowledge. Cf. Aronson 1989: 231-7; Chihara 1981: 431-2, 451 n. 35; Clark 2002: 164; Earman 1992: 70; Good 1968; Hempel 1967; Horwich 1982: 62; Howson & Urbach 1993: 128-9; Korb 1994; Maher 1999: 62-4; Poundstone 1988: 39-40; Rosenkrantz 1977: 33-5, 1982: 82-4; Sainsbury 1995: 81; Swinburne 1971: 326-7; Will 1966: 56-7.

The standard Bayesian solution answers the prescriptive question with “yes, but marginally”, and answers the explanatory question with the claim that people mistake marginal confirmation for confirmational irrelevance. If the disputed assumption is false, then Bayesians must give up the above answer to the prescriptive question: $P(H|\overline{Ba}\overline{Ra})$ may be equal to or different from $P(H)$, and which of these cases should hold depends (approximately) on whether $P(Ba|H)$ is equal to or different from $P(Ba)$.²⁵ But Bayesians can still produce a plausible answer to the explanatory question: they can supplement their previous answer to this question with the claim that *people mistakenly take the disputed assumption to be true*. Why do people make this mistake? Because, one might suggest, they reason by *indifference* (see §3). Take a standard example (van Fraassen 1989: 303): given that a cube comes from a factory which produces cubes with edge length at most 2 cm, what is the probability that the cube has edge length at most 1 cm? Most people, one might conjecture, would say it’s 1/2. Similarly, one might conjecture, most people would say that $P(Ba|H)$ and $P(Ba|\overline{H})$ should be equal. Reasoning by indifference comes naturally to people; it takes some thought to realize that such reasoning leads to inconsistencies.²⁶

Bayesians who want to pursue the above suggestion have a lot of work to do: given that the explanatory question is empirical, they need to go beyond considerations of plausibility and examine evidence on how people in fact reason (cf. Humberstone 1994). So I hope it is clear that the above suggestion is not intended as a full defense of a Bayesian solution to Hempel’s paradox. (Such a defense would also need to address the numerous objections that have been raised against the standard Bayesian solution.) The suggestion has instead the limited but useful purpose of showing how Bayesians might be able to cope with a little noticed but particularly perfidious obstacle: the implausibility of the disputed assumption.

6. Conclusion.

The disputed assumption, namely the claim that $P(Ba|H)$ and $P(Ba)$ should be equal, is both sufficient and for all practical purposes necessary for the standard Bayesian claim that $P(H|\overline{Ba}\overline{Ra}) - P(H)$ should be positive but minute; i.e., the claim that $\overline{Ba}\overline{Ra}$ does confirm H but only marginally. I argued that the assumption is implausible: $P(Ba|H)$ and $P(Ba)$ need not be equal. This is not to say that they *should* differ: they may differ but they may also be equal. If so, then the standard Bayesian answer to the *prescriptive* question must be given up: $P(H|\overline{Ba}\overline{Ra})$ and $P(H)$ may differ but they may also be equal. Bayesians, however, can still produce a plausible answer to the *explanatory* question: people, reasoning by indifference, mistakenly take the disputed assumption to be true, and then mistake marginal confirmation for confirmational irrelevance. I conclude that, even if the assumption is false, there is still hope for Bayesians.

Appendix: The Principle of Conditional Indifference leads to inconsistency.

Suppose that an object a is drawn by simple random sampling from a finite population. Let the random variables X , Y , and Z represent respectively the percentages of ravens, of black objects, and of black ravens in

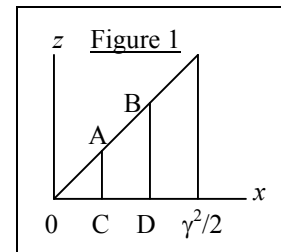
²⁵ If so, and if the Equivalence Condition is true, then Bayesians must also give up Nicod’s Condition (see §2).

²⁶ To reach an inconsistency in the factory example, note that the factory produces cubes of volume at most 8 cm³, and we are asking what is the probability that a cube from the factory has volume at most 1 cm³; so why shouldn’t the answer be 1/8 rather than 1/2? To reach an inconsistency in the context of the disputed assumption, see the Appendix.

the population (so that the percentage of ravens that are black is $G = Z/X$). Define the (Bernoulli) random variable R to have the value 1 if a is a raven and 0 otherwise. Then $P(R=1|X=x)$ should be x . Given that the population contains overwhelmingly more nonblack objects than ravens, the space of all possible values for $\langle X, Y, Z \rangle$ is $S = \{\langle x, y, z \rangle: 0 \leq z \leq x \leq \gamma^2(1-y) \wedge z \leq y\}$, with γ^2 positive but minute (ignoring, to simplify, the complication that $x, y,$ and z can take only discrete values). To simplify the exposition, I will deal with the intersection S' of S with the plane (e.g.) $y = 1/2$. S' corresponds to the largest triangular area in Figure 1, the hypotenuse corresponding to H .

The key to generating an inconsistency by means of PCI is that a sample space can be partitioned in multiple ways. The family of lines $\{z = gx; 0 \leq g \leq 1\}$ gives one partition of S' ; the family of lines $\{z = x^n; n \geq 1\}$ (plus the line $z = 0$) gives another partition. (Strictly speaking, these are not partitions, because every line includes the point $\langle 0, 0 \rangle$; but PCI is not limited to partitions.) Since a given value g of G is compatible with all possible values x of X , g provides no information on the distribution of R (which is determined by x); so PCI entails that, for any g and g' , $P(R=1|G=g)$ and $P(R=1|G=g')$ should be equal to each other—and thus also to $P(R=1)$. It follows that R and $G = f_1(X, Z) = Z/X$ should be independent. Similarly, PCI entails that R and $N = f_2(X, Z) = \ln Z / \ln X$ (this comes from $Z = X^N$) should be independent. In the end, PCI will entail that R and $\langle X, Z \rangle$ should be independent, a conclusion incompatible with the claim that $P(R=1|X=x)$ should be x .

To derive an inconsistency rigorously, consider the simplified case in which exactly four points of S' have nonzero probabilities (Figure 1). (Such a prior does not seem epistemically irrational, unless one appeals to the demonstrably problematic PI.) Assume that the probabilities $p_A, p_B, p_C,$ and p_D are all different and sum to 1, and that the x -coordinates satisfy $x_B = x_D$ and $x_A = x_C$. The last equality gives (in obvious simplified notation; e.g., $P(R|A) = P(R=1|X=x_A \wedge Y=1/2 \wedge Z=z_A)$) $P(R|A) = P(R|C)$, so that (using Bayes' theorem) (1) $p_C P(A|R) = p_A P(C|R)$. Similarly, $x_B = x_D$ gives: (2) $p_D P(B|R) = p_B P(D|R)$. Applying now PCI to the partition $\{A \vee B, C \vee D\}$ gives that two equalities should hold: $P(R|A \vee B) = P(R) = P(R|C \vee D)$. The first equality gives $P(A \vee B|R) = P(A \vee B)$, so (3) $P(A|R) + P(B|R) = p_A + p_B$. The second equality gives similarly: (4) $P(C|R) + P(D|R) = p_C + p_D$. A similar application of PCI to the partition $\{A \vee D, C \vee B\}$ gives: (5) $P(A|R) + P(D|R) = p_A + p_D$, (6) $P(C|R) + P(B|R) = p_C + p_B$. It can be shown that the system of equations (1) - (6) has the unique solution $P(A|R) = p_A, P(B|R) = p_B, P(C|R) = p_C,$ and $P(D|R) = p_D$. But then $P(R|A) = P(A|R)P(R)/p_A = P(R) = P(B|R)P(R)/p_B = P(R|B)$. The conclusion that $P(R|A)$ and $P(R|B)$ should be equal is incompatible with $x_A \neq x_B$.



REFERENCES

- Alexander, H. G. (1958), "The paradoxes of confirmation", *British Journal for the Philosophy of Science* 9: 227-233.
- Aronson, Jerold R. (1989), "The Bayesians and the raven paradox", *Noûs* 23: 221-240.
- Black, Max (1966), "Notes on the 'paradoxes of confirmation'", in Jaakko Hintikka and Patrick Suppes (eds.), *Aspects of inductive logic*. Amsterdam: North-Holland, 175-197.
- Chihara, Charles (1981), "Quine and the confirmational paradoxes", in Peter A. French, Theodore E. Uehling Jr., and Howard K. Wettstein (eds.), *Midwest Studies in Philosophy. Vol. 6: The foundations of analytic philosophy*. Minneapolis: University of Minnesota Press, 425-452.
- Christensen, David (1999), "Measuring confirmation", *Journal of Philosophy* 96: 437-461.
- Clark, Michael (2002), *Paradoxes from A to Z*. New York: Routledge.
- Cohen, Yael (1987), "Ravens and relevance", *Erkenntnis* 26: 153-179.
- Earman, John (1992), *Bayes or bust? A critical examination of Bayesian confirmation theory*. Cambridge, MA: MIT Press.
- Eells, Ellery (1982), *Rational decision and causality*. New York: Cambridge University Press.
- Eells, Ellery and Branden Fitelson (2000), "Measuring confirmation and evidence", *Journal of Philosophy* 97: 663-672.
- Fisch, Menachem (1984), "Hempel's ravens, the natural classification of hypotheses and the growth of knowledge", *Erkenntnis* 21: 45-62.
- Fitelson, Branden (1999), "The plurality of Bayesian measures of confirmation and the problem of measure sensitivity", *Philosophy of Science* 66 (Proceedings): S362-S378.
- (2001), "A Bayesian account of independent evidence with applications", *Philosophy of Science* 68 (Proceedings): S123-S140.

- Forster, Malcolm R. (1994), "Non-Bayesian foundations for statistical estimation, prediction, and the ravens example", *Erkenntnis* 40: 357-376.
- Foster, Lawrence (1971), "Hempel, Scheffler, and the ravens", *Journal of Philosophy* 68: 107-114.
- Gaifman, Haim (1979), "Subjective probability, natural predicates and Hempel's ravens", *Erkenntnis* 14: 105-147.
- Gibson, L. (1969), "On 'ravens and relevance' and a likelihood solution of the paradox of confirmation", *British Journal for the Philosophy of Science* 20: 75-80.
- Giere, Ronald N. (1970), "An orthodox statistical resolution of the paradox of confirmation", *Philosophy of Science* 37: 354-362.
- Good, Irving J. (1960), "The paradox of confirmation", *British Journal for the Philosophy of Science* 11: 145-149.
- (1961), "The paradox of confirmation (II)", *British Journal for the Philosophy of Science* 12: 63-64.
- (1967), "The white shoe is a red herring", *British Journal for the Philosophy of Science* 17: 322.
- (1968), "The white shoe *qua* herring is pink", *British Journal for the Philosophy of Science* 19: 156-157.
- Goodman, Nelson (1983), *Fact, fiction, and forecast* (4th ed.). Cambridge, MA: Harvard University Press.
- Grandy, Richard E. (1967), "Some comments on confirmation and selective confirmation", *Philosophical Studies* 18: 19-24.
- Harré, R. (1970), *The principles of scientific thinking*. London: MacMillan.
- Hempel, Carl G. (1937), "Le problème de la vérité", *Theoria* 3: 206-246.
- (1943), "A purely syntactical definition of confirmation", *Journal of Symbolic Logic* 8: 122-143.
- (1945), "Studies in the logic of confirmation", *Mind* 54: 1-26, 97-121.
- (1965), *Aspects of scientific explanation and other essays in the philosophy of science*. New York: Free Press.
- (1967), "The white shoe: No red herring", *British Journal for the Philosophy of Science* 18: 239-240.
- Hesse, Mary (1974), *The structure of scientific inference*. London: MacMillan.
- Hintikka, Jaakko (1969), "Inductive independence and the paradoxes of confirmation", in Nicholas Rescher (ed.), *Essays in honor of Carl G. Hempel: A tribute on the occasion of his sixty-fifth birthday*. Dordrecht: Reidel, 24-46.
- Hooker, Clifford A. and D. Stove (1968), "Relevance and the ravens", *British Journal for the Philosophy of Science* 18: 305-315.
- Horwich, Paul (1982), *Probability and evidence*. New York: Cambridge University Press.
- Hosiasson-Lindenbaum, Janina (1940), "On confirmation", *Journal of Symbolic Logic* 5: 133-148.
- Howson, Colin and Peter Urbach (1993), *Scientific reasoning: The Bayesian approach* (2nd ed.). Chicago: Open Court.
- Huggett, W. J. (1960), "On not being gulled by ravens", *Australasian Journal of Philosophy* 38: 48-50.
- Humberstone, I. L. (1994), "Hempel meets Wason", *Erkenntnis* 41: 391-402.
- Humburg, Jürgen (1986), "The solution of Hempel's raven paradox in Rudolf Carnap's system of inductive logic", *Erkenntnis* 24: 57-72.
- Jardine, R. (1965), "The resolution of the confirmation paradox", *Australasian Journal of Philosophy* 43: 359-368.
- Jeffrey, Richard C. (1995), "A brief guide to the work of Carl Gustav Hempel", *Erkenntnis* 42: 3-7.
- Joyce, James M. (1999), *The foundations of causal decision theory*. New York: Cambridge University Press.
- Korb, Kevin B. (1994), "Infinitely many resolutions of Hempel's paradox", in R. Fagin (ed.), *Theoretical aspects of reasoning about knowledge: Proceedings of the fifth conference*. Asilomar, CA: Morgan Kaufmann, 138-149.
- Kruse, Michael (2000), "Are there Bayesian success stories? The case of the ravens paradox", presented at the 2000 Philosophy of Science Association conference. Available at <<http://hypatia.ss.uci.edu/lps/psa2k/bayesian-success-stories.pdf>>.
- Lewis, David K. (1980), "A subjectivist's guide to objective chance", in Richard C. Jeffrey (ed.), *Studies in inductive logic and probability*, vol. 2. Berkeley: University of California Press, 263-294.
- (1994), "Humean Supervenience debugged", *Mind* 103: 473-490.
- Lipton, Peter (1991), *Inference to the best explanation*. New York: Routledge.
- Mackie, John L. (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.
- (2002), "Kuipers on qualitative confirmation and the ravens paradox", forthcoming in Roberto Festa et al. (eds.), *Confirmation, empirical progress, and truth approximation: Essays in debate with Theo Kuipers*, vol. 1. Amsterdam: Rodopi.
- Morgenbesser, Sidney (1962), "Goodman on the ravens", *Journal of Philosophy* 59: 493-495.
- Nerlich, G. (1964), "Mr. Wilson on the paradox of confirmation", *Australasian Journal of Philosophy* 42: 401-405.
- Nicod, Jean (1924), *Le problème logique de l'induction*. Paris: Alcan.
- (1930), *Foundations of geometry and induction* (P. P. Wiener, trans.). London: Routledge.
- Pennock, Robert T. (1991), *Causally licensed inference and the confirmation relation*. Ph. D. dissertation, University of Pittsburgh.

- Poundstone, William (1988), *Labyrinths of reason: Paradox, puzzles, and the frailty of knowledge*. New York: Doubleday.
- Rescher, Nicholas (2001), *Paradoxes: Their roots, range, and resolution*. Chicago: Open Court.
- Rody, Phillip J. (1978), "(C) instances, the relevance criterion, and the paradoxes of confirmation", *Philosophy of Science* 45: 289-302.
- Rosenkrantz, Roger D. (1977), *Inference, method and decision*. Boston: Reidel.
- (1982), "Does the philosophy of induction rest on a mistake?", *Journal of Philosophy* 79: 78-97.
- Royall, Richard M. (1997), *Statistical evidence: A likelihood paradigm*. New York: Chapman & Hall.
- Sainsbury, Richard M. (1995), *Paradoxes* (2nd ed.). New York: Cambridge University Press.
- Scheffler, Israel (1963), *The anatomy of inquiry: Philosophical studies in the theory of science*. New York: Knopf.
- Scheffler, Israel and Nelson Goodman (1972), "Selective confirmation and the ravens: A reply to Foster", *Journal of Philosophy* 69: 78-83.
- Schoenberg, Judith (1964), "Confirmation by observation and the paradox of the ravens", *British Journal for the Philosophy of Science* 15: 200-212.
- Schwartz, Robert (1972), "Paradox and projection", *Philosophy of Science* 39: 245-248.
- Skyrms, Brian (1966), "Nomological necessity and the paradoxes of confirmation", *Philosophy of Science* 33: 230-249.
- Smokler, Howard E. (1967), "The equivalence condition", *American Philosophical Quarterly* 4: 300-307.
- Stillwell, Shelley (1985), "Confirmation, paradoxes, and possible worlds", *British Journal for the Philosophy of Science* 36: 19-52.
- Stove, D. (1966), "Hempel's paradox", *Dialogue* 4: 444-455.
- Suppes, Patrick (1966), "A Bayesian approach to the paradoxes of confirmation", in Jaakko Hintikka and Patrick Suppes (eds.), *Aspects of inductive logic*. Amsterdam: North-Holland, 198-207.
- Swinburne, Richard G. (1971), "The paradoxes of confirmation—A survey", *American Philosophical Quarterly* 8: 318-330.
- (1973), *An introduction to confirmation theory*. London: Methuen.
- Tempelmeier, Ernest S. (1972), *The raven paradoxes and epistemological problems of confirmation theory*. Ph. D. dissertation, Temple University.
- Tuske, Joerg (1998), "Dinnāga and the raven paradox", *Journal of Indian Philosophy* 26: 387-403.
- van Fraassen, Bas C. (1984), "Belief and the will", *Journal of Philosophy* 81: 235-256.
- (1989), *Laws and symmetry*. Oxford: Clarendon.
- (1995), "Belief and the problem of Ulysses and the sirens", *Philosophical Studies* 77: 7-37.
- Vincent, R. H. (1975), "Selective confirmation and the ravens", *Dialogue* 14: 3-49.
- von Wright, Georg H. (1966), "The paradoxes of confirmation", in Jaakko Hintikka and Patrick Suppes (eds.), *Aspects of inductive logic*. Amsterdam: North-Holland, 208-218.
- Will, Frederick L. (1966), "Consequences and confirmation", *Philosophical Review* 75: 34-58.
- Wilson, P. R. (1964a), "On the confirmation paradox", *British Journal for the Philosophy of Science* 15: 196-199.
- (1964b), "A new approach to the confirmation paradox", *Australasian Journal of Philosophy* 42: 393-401.
- Woodward, James (1985), "Critical review: Horwich on the ravens, projectability and induction", *Philosophical Studies* 47: 409-428.
- Young, Gregory A. (1975), *The paradoxes of confirmation: A study in the logic of confirmation*. Ph. D. dissertation, The University of Miami.