

The Wason Task(s) & The Paradox of Confirmation

Branden Fitelson

Department of Philosophy
Group in Logic and the Methodology of Science
&
Cognitive Sciences Core Faculty
University of California-Berkeley

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

- **Nicod Condition (NC):** For any object x and any properties ϕ and ψ , the proposition that x is both ϕ and ψ confirms the proposition that every ϕ is ψ . 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*) logically equivalent to H_2 , then E confirms H_2 . More formally:

$$\text{If } E \text{ confirms } H_1, \text{ and } H_1 \models H_2, \text{ then } E \text{ confirms } H_2.$$

- **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 [8] & Goodman [7] *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 $(\forall x)(Rx \supset Bx)$ [12].]

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 .

☞ Hempel's *theory entails* (M) [4]. Good intuition [(I)], bad theory.

Unlike Hempel, Bayesians (*e.g.*, Carnap [1]) use *probabilistic relevance* relations to explicate the confirmation relation.

This has several advantages over Hempel's *deductive* approach:

- 1 It leads to a *non-monotonic* confirmation relation, which can accommodate Hempelian *anti*-(M) *intuitions* like (I).
- 2 It gives rise to a confirmation relation which does *not* imply (NC). [See "Extras" and [13] for examples and discussion.]
- 3 It supplies *comparative* (and quantitative) c -relations:
 - E_1 confirms H *more strongly than* E_2 does — relative to background corpus K — iff $\text{Pr}(H | E_1 \ \& \ K) > \text{Pr}(H | E_2 \ \& \ K)$.
[$c(H, E | K) \stackrel{\text{def}}{=} \text{the degree to which } E \text{ confirms } H \text{ (rel. to } K).$]

Next, a brief review of the canonical comparative Bayesian response(s) to The Paradox. Then, it's on to Wason's Task(s).

There have been *many* comparative Bayesian approaches to the paradox (see [19]). Here is a canonical characterization:

Assume that our *actual* background corpus K_α is such that:

- (4) $\Pr(\sim Ba \mid K_\alpha) > \Pr(Ra \mid K_\alpha)$
- (5) $\Pr(Ra \mid H \& K_\alpha) = \Pr(Ra \mid K_\alpha)$ [$\therefore \Pr(\sim Ra \mid H \& K_\alpha) = \Pr(\sim Ra \mid K_\alpha)$!]
- (6) $\Pr(\sim Ba \mid H \& K_\alpha) = \Pr(\sim Ba \mid K_\alpha)$ [$\therefore \Pr(Ba \mid H \& K_\alpha) = \Pr(Ba \mid K_\alpha)$!]

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

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

\therefore the proposition that *a* is a black raven (*actually*) confirms that all ravens are black *more strongly than* the proposition that *a* is a nonblack nonraven, *if* (4)–(6) hold for (*actual*) K_α .

(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, just note that Hempel, Carnap, *et al.* would reject 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 [19] 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 [19] discusses (5) & (6), and their standard rationales.

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

- (‡) $\Pr(H \mid Ra \& K_\alpha) \approx \Pr(H \mid \sim Ba \& K_\alpha)$

☞ (4) & (‡) imply (7) — no independence assumptions needed [4].

(‡) says: Ra confirms H to \approx the same degree as $\sim Ba$ does. This assumption is far more plausible than the independencies (5) & (6). None of the standard arguments against (5)/(6) apply to (‡).

Moreover, accepting (4) & (‡) 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.

Hempel's own line on the paradox favors (4) & (‡), which is compatible with (PC) & \neg (PC*). Interestingly, assumptions (4)–(6) are *not* compatible with Hempel's line (nor Carnap's [12, 13])!

- Wason gives various versions of his “task(s)”. *E.g.*,
Given the sentence: Every card which has a D on one side has a 3 on the other side (and knowledge that each card has a letter on one side and a number on the other), together with four cards showing D, K, 3, 7, hardly any individuals make the correct choice of cards to turn over (D, 7) in order to determine the truth of the sentence. [20, p. 63]
- This characterization is unclear. Here is a precisification:
Each card (in some set of cards C) has one letter on one side and one number on the other side. You will be shown four cards from C (with one face down), and you will be asked to turn over one or more of the four cards, with an eye toward determining whether the following hypothesis is true:

(H) All “D”-cards (in C) are “3”-cards.

Q: Which of the following 4 cards would you turn to test H?

D	K	3	7
---	---	---	---

- Empirically, the most frequent answers are (in decreasing order of f): (i)

D	3
---	---

, (ii)

D

, (iii)

D	3	7
---	---	---

, (iv)

D	7
---	---

; and, for single cards, the frequencies are ordered:

D

 >

3

 >

7

.

- Humberstone [10] was the first to draw an explicit analogy between Wason’s task(s) and the Paradox of Confirmation.
- Unfortunately, Humberstone seems not to have been read by the cognitive scientists who (later) exploit the analogy.
- Humberstone’s key insight is that Wason’s original description—which *leaves out C*—was *ambiguous*.
- Wason presupposes (without telling his subjects!) that C just *is* the set of four cards you are shown. Two readings:
 - (1) $C = \boxed{D} \boxed{K} \boxed{3} \boxed{7}$. This is the reading Wason presupposes.
 - (2) $\boxed{D} \boxed{K} \boxed{3} \boxed{7} \subset C$. This is Humberstone’s alternative.
- As Wason notes, (1) implies that there is a *single, definitive, correct answer* to the Question: $E_w = \boxed{D} \boxed{7}$ — choice (iv)!
- Wason must have had (1) in mind, since he talks as if E_w is *the answer*. But, (2) is also consistent with his descriptions.
- (2) leads to an analogy with the Paradox of Confirmation. Humberstone sketches this analogy, in a *Hempel*ian way.

- Recall: Hempel’s Paradox was meant to involve examining objects that you — antecedently — know *nothing* about [remember the role of Hempel’s intuitive (PC)/(PC*) distinction].
- This is *disanalogous* with the Wason Task, since you get to see one side of the cards you’ll be examining *in advance*.
- To shore-up this disanalogy, consider the following (only slightly modified) “Wason-style” confirmation task:
 You will be testing a hypothesis about a set of birds **B**. For each bird in **B**, we will make a card (in **C**), which will have R or $\sim R$ (for raven/non-raven) on one side, and B or $\sim B$ (black/non-black) on the other. You will be shown the cards of four birds in **B** (with one face down), and you will be asked to turn over one or more of the four cards, with an eye toward testing the following:
 (H') All “ R ”s (in **C**) are “ B ”s. [*i.e.*, All ravens (in **B**) are black.]
 You’re shown these 4 cards from **C** [part of your *background K*]:
 $\boxed{R} \boxed{\sim R} \boxed{B} \boxed{\sim B}$
 Q' : Which of these card(s) would you turn over to test H' ?

- The analogous empirical ordering is: (i) $\boxed{R} \boxed{B}$, (ii) $\boxed{R} \boxed{\sim B}$, (iii) $\boxed{R} \boxed{B} \boxed{\sim B}$, (iv) $\boxed{R} \boxed{\sim B}$ [and, for single cards: $\boxed{R} > \boxed{B} > \boxed{\sim B}$].
- And, as before, if $C = \boxed{R} \boxed{\sim R} \boxed{B} \boxed{\sim B}$, then $E_w = \boxed{R} \boxed{\sim B}$ (iv) is *the* correct answer. But, what if $\boxed{R} \boxed{\sim R} \boxed{B} \boxed{\sim B} \subset C$?
- Now, (Q') is *similar* to The Paradox, on a *two-stage sampling* model of evidence-gathering (*e.g.*, [18]). Three strategies:
 - (a) Sampling an object a from the class of ravens and then checking to see whether a is black. [turning over \boxed{R}]
 - (b) Sampling an object a from the class of black things and checking to see whether a is a raven. [turning over \boxed{B}]
 - (c) Sampling an object a from the class of non-black things and checking to see whether a is a raven. [turning over $\boxed{\sim B}$]
- The Bayesians’ (4)–(6) imply that (a) generates better evidence than (c) — *if* both generate *confirmatory* evidence.
- But, since $|c(H, E | K)| \sim \square = |c(H, \sim E | K)|$ [3], this doesn’t explain why (a) should yield better evidence “on average”.
- Also, (4)–(6) don’t have implications for (b)’s place in an ordering of the “expected quality of generated evidence”.

- Nickerson [15] exploits the two-stage Bayesian approaches to The (modified) Paradox, to “rationalize” or “justify” the empirical responses, which Wason thought were “defective”.
- Nickerson’s approach involves the adoption of the following simple measure of *absolute* degree of confirmation:
 $\bar{d}(H, E | K) \stackrel{\text{def}}{=} |\Pr(H | E \& K) - \Pr(H | K)|$
- Nickerson uses \bar{d} to define the “expected confirmational utility” [$u(S)$] of an evidence-gathering strategy (S).
- He does this for single-card-turning strategies in our (modified) Paradox of Confirmation set-up, as follows:
 $u(\boxed{R}) \stackrel{\text{def}}{=} \Pr(Ba | Ra) \cdot \bar{d}(H', Ba | Ra) + \Pr(\sim Ba | Ra) \cdot \bar{d}(H', \sim Ba | Ra)$
 $u(\boxed{\sim R}) \stackrel{\text{def}}{=} \Pr(Ba | \sim Ra) \cdot \bar{d}(H', Ba | \sim Ra) + \Pr(\sim Ba | \sim Ra) \cdot \bar{d}(H', \sim Ba | \sim Ra)$
 $u(\boxed{B}) \stackrel{\text{def}}{=} \Pr(Ra | Ba) \cdot \bar{d}(H', Ra | Ba) + \Pr(\sim Ra | Ba) \cdot \bar{d}(H', \sim Ra | Ba)$
 $u(\boxed{\sim B}) \stackrel{\text{def}}{=} \Pr(Ra | \sim Ba) \cdot \bar{d}(H', Ra | \sim Ba) + \Pr(\sim Ra | \sim Ba) \cdot \bar{d}(H', \sim Ra | \sim Ba)$
- Then, he writes down a (precise, *numerical*) distribution \Pr_n , which *entails* the standard Bayesian (4)–(6), *and* which yields a u -ordering that agrees with the empirical: $\boxed{R} > \boxed{B} > \boxed{\sim B}$.

- Nickerson's ordering (○) $u(\boxed{R}) > u(\boxed{B}) > u(\boxed{\sim B})$ is *not entailed* by (4)–(6). However, (4')–(6) *do* entail (○), where: (4') $\Pr(\sim Ba) > \Pr(Ba) > \Pr(Ra)$
- This gives simple sufficient conditions for Nickerson's ordering (he does not mention this; he only reports \Pr_n).
- As we have seen, (4')–(6) have undesirable consequences. But, (4')–(6) are *not necessary* for (○). More precise facts:
- (4) & (‡) $\Rightarrow u(\boxed{R}) > u(\boxed{\sim B})$. But, (4') & (‡) $\not\Rightarrow$ (○). Adding (★) $\Pr(H | R) = \Pr(H | B)$ to (4') & (‡) suffices, but \Rightarrow (4')–(6).
 - What is the weakest condition (C) s.t. (4') & (‡) & (C) \Rightarrow (○)?
- (4') & (‡) & (○) *contradicts* this triple [but (4') & (‡) does *not*]:
 - $\Pr(H | Ra \& Ba) \geq \Pr(H | Ba)$.
 - $\Pr(H) \geq \frac{1}{2}$.
 - $\Pr(Ra | Ba) \leq \Pr(Ra)$.
- (○) *by itself contradicts* Hempelian and Carnapian theories of ϵ , since all such (inductive-logical) theories entail that $\epsilon(H', Ra | Ba) = \epsilon(H', \sim Ra | Ba) = 0$.

- [1] R. Carnap, *Logical Foundations of Probability*, 2nd ed., Chicago U. Press, 1962.
- [2] L. Cosmides, *The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task*, *Cognition* 31 (1985), 187-276.
- [3] E. Eells and B. Fitelson, *Symmetries and Asymmetries in Evidential Support*, *Philosophical Studies*, 107 (2002), 129-142.
- [4] B. Fitelson and J. Hawthorne, *How Bayesian confirmation theory handles the paradox of the ravens*, *Probability in Science* (Eells and Fetzer, eds.), to appear (2009).
- [5] I.J. Good, *The white shoe is a red herring*, *British J. for the Phil. of Sci.* 17 (1967), 322.
- [6] ———, *The white shoe qua red herring is pink*, *British J. for the Phil. of Sci.*, 19 (1968), 156-7.
- [7] N. Goodman, *Fact, Fiction, and Forecast*, Harvard University Press, 1955.
- [8] C. Hempel, *Studies in the logic of confirmation*, *Mind* 54 (1945), 1-26, 97-121.
- [9] ———, *The white shoe: no red herring*, *British J. for the Phil. of Sci.* 18 (1967), 239-240.
- [10] Humberstone, L., *Hempel Meets Wason*, *Erkenntnis* 41 (1994), 391-402.
- [11] J. MacFarlane and N. Kolodny, *Ifs and Oughts*, unpublished manuscript (2008).
- [12] P. Maher, *Inductive logic and the ravens paradox*, *Philosophy of Science* 66 (1999), 50-70.
- [13] ———, *Probability captures the logic of scientific confirmation*, *Contemporary Debates in the Philosophy of Science* (Christopher Hitchcock, ed.), Blackwell, 2004.
- [14] C. McKenzie and L. Mikkelsen, *The Psychological Side of Hempel's Paradox of Confirmation*, *Psychonomic Bulletin & Review* 7 (2000), 360-66.
- [15] Nickerson, R. *Hempel's Paradox and Wason's Selection Task: Logical and Psychological Puzzles of Confirmation*, *Thinking and Reasoning* 2 (1996), 1-31.
- [16] Oaksford, M. and Chater N. *A Rational Analysis of the Selection Task as Optimal Data Selection*, *Psychological Review* 101 (1994), 608-631.
- [17] W.V.O. Quine, *Natural kinds, Ontological Relativity and Other Essays*, Columbia U. Press, 1969.
- [18] Royall, R. *Statistical Evidence: A Likelihood Paradigm*, CRC Press, 1999.
- [19] P. Vranas, *Hempel's raven paradox: a lacuna in the standard Bayesian solution*, *British Journal for the Philosophy of Science* 55 (2004), 545-560.
- [20] Wason, P. and D. Shapiro. *Natural and Contrived Experience in a Reasoning Problem*, *Quarterly Journal of Experimental Psychology* 23 (1971), 63-71.

I.J. Good [5] gave the following Bayesian counterexample to (NC):

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:

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

$\therefore E$ lowers the probability of (disconfirms) H , relative to K .

\therefore (NC) is false, and *even for "natural kinds"* (pace Quine [17]). Similar examples can be used to show that (PC) is also false.

Hempel [9] complains that Good's example is not probative, since (NC) must be taken relative to *empty background* $K = \top$.

Is this a fair complaint? [No — (M)!] Anyhow, Good responds ...

Here's Good's [6] attempt to meet Hempel's $K = \top$ 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 confident about this $K = \top$ counterexample. Maher [12] argues this is not a compelling counterexample.

Maher [13] has recently provided a more compelling (*Carnapian*) counterexample to (NC), which is beyond our scope today.¹

Most Bayesians don't *understand* $(NC_{K=\top})$. Unlike Carnap [1], they have *no theory* of " \Pr_{\top} " [or " $\mathcal{C}(H, E | \top)$ "]. So, they opt for a different sort of approach, using *epistemic* \Pr and *actual* $K = K_{\alpha}$.

¹Maher [13] shows that $\Pr_{\top}(H | E) < \Pr_{\top}(H)$, for some adequate Carnapian \Pr_{\top} functions. Hence, (NC) is false for a Carnapian theory of " $\mathcal{C}(H, E | \top)$ ".

McKenzie & Mikkelsen (M&M) [14] report Ψ -experiments involving a variety of “Hempel-like” hypothesis-testing problems.

Their data show that changes in the “rarity assumption” [(4')] are correlated with changes in agents' responses as to the degree to which $(E_2) \sim Xa \& \sim Ya$ is comparatively probative [vs $(E_1) Xa \& Ya$], concerning (H) All X 's are Y 's (for *many* X 's and Y 's).

M&M see this as “normative”, since their normative model makes similar predictions. I have three comments on their model:

- Like Nickerson & typical Bayesians, M&M assume (5) & (6).
- Unlike Hempel/Bayesians who assume agents test H against $\sim H$, M&M suppose that agents test H against H' , where H' asserts that $X \perp\!\!\!\perp Y$. This is a “Likelihoodist” approach [18].
- M&M *try* to draw the Hempel/Wason analogy, but they seem insensitive to the fact that explaining the *Wason* data requires explaining $\boxed{R} > \boxed{B} > \boxed{\sim B}$, and *not merely* $\boxed{R} > \boxed{\sim B}$. [\therefore some “advantages” they claim for their model are *misleading*.]

- Oaksford & Chater [16] give yet another “rationalization” (Bayesian-style) of the responses to Wason’s Task(s).
- In some respects, their approach is similar to that of M & M:
 - O & C do not test H' against $\sim H'$ (or, “in isolation”, as they put it). Rather, they test H' against H'' , which is the hypothesis that R and B are *probabilistically independent*.
 - Like M&M, this is more of a “Likelihoodist” [18] approach.
- In other respects, O&C’s approach is similar to Nickerson’s:
 - O & C define their “ u ” in terms of *expected information gain* (or expected *entropy decrease*), which is more similar (than M&M is) to Nickerson’s *expected degree of confirmation*.
- In still other respects, their approach is dissimilar to all other Bayesian approaches, as they do *not* assume *independencies* (5)/(6), or even our [4] weaker (\ddagger).
 - They (now) seem to think their account is more similar to Nickerson’s. I need to examine their models more closely before rendering an opinion. But, if they are like Nickerson, they should inherit some of his problems (explained above).

Cosmides [2] reports “Wason-like” experiments in which agents seem to do “better” — *assuming Wason’s normative model*.

Her examples involve conditionals with normative and/or modal content in their consequents. *E.g.*, she asks subjects to test:

If a person is drinking beer (D), then he must be over 20 years old (O).

by turning one or more of these 4 cards [where one side has a person’s drinking behavior $D/\sim D$ and the other has their age $O/\sim O$]:

$\boxed{D} \quad \boxed{\sim D} \quad \boxed{O} \quad \boxed{\sim O}$

The data for Cosmides’s “Wason-like” tasks fit Wason’s normative $\boxed{D} \geq \boxed{\sim O} > \boxed{O}$ much better than Wason’s data did.

Cosmides thinks this is “*good news*” for actual subjects, and evidence that evolution has made us “better” at testing certain types of normatively/modally loaded hypotheses/conditionals.

☞ Recent work in the semantics of such conditionals [11] suggests contraposition is *invalid* for them! Is [2] *Grist for Wason’s Mill*?