Probability, Confirmation, and the "Conjunction Fallacy"

Branden Fitelson[1]

Department of Philosophy & Religion

Northeastern University

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

[1] This is based on joint work with Vincenzo Crupi (Turin) & Katya Tentori (Trento).

1. Overview
2. Some Background on Probability (Blackboard)
3. Historical Background I: Carnapian Confirmation Theory
   - Confirmation, "Logical" Probability, and Relevance
4. Historical Background II: Bayesian Confirmation Theory
   - Confirmation and Subjective Probabilistic Relevance
5. Applying Bayesian Confirmation to the "Conjunction Fallacy"
   - The Traditional "Conjunction Fallacy" Cases
   - A Confirmation-Theoretic Approach to the Traditional CFs
   - Non-Traditional CFs and Bayesian Confirmation Theory

Hempel

Carnap

Popper

- In the first edition of LFP, Carnap [3] undertakes a precise probabilistic explication of the concept of confirmation. This is where modern confirmation theory was born (in sin).
- Carnap was interested mainly in *quantitative* confirmation (which he took to be fundamental). But, he also gave (derivative) qualitative and comparative explications:
  - Qualitative. E inductively supports H.
  - **Comparative**. E supports H *more strongly than* E' supports H'.
  - Quantitative. E inductively supports H *to degree* r.
- Carnap begins by clarifying the *explicandum* (the informal "inductive support" concept) in various ways, including:
  - Qualitative. $(\star)$ E gives some (positive) evidence for H.
- Note two things. First, $(\star)$ sounds *epistemic* (not *logical*). Second, $(\star)$ sounds like it involves (positive) *relevance*.
- Strangely, Carnap proceeds (in LFP_1) to offer a *logical* account of confirmation that does *not* involve relevance.
- These were the two original sins of Bayesian confirmation...

- In the 1st ed. of LFP, Carnap characterizes "the degree to which E confirms H" as $\mathfrak{c}(H,E) = \Pr(H \mid E)$, which leads to:

  Quantitative. $\Pr(H \mid E) = r$.

  Comparative. $\Pr(H \mid E) > \Pr(H' \mid E')$.

  Qualitative. $\Pr(H \mid E) > t$ (typically, with "threshold" $t > \frac{1}{2}$).

  - Doesn't sound like ($\star$). More on this dissonance below.
- Like Hempel [8], Carnap wanted a *logical* explication of confirmation (as a relation between sentences in FO$\mathcal{L}$s).
- For Carnap, this meant that the probability functions used in confirmation theory must *themselves* be "logical".
- This leads naturally to the Carnapian project of providing a "logical explication" of conditional probability $\Pr(\cdot \mid \cdot)$ *itself*.
- Here, Carnap was strongly influenced by Keynes [10], who believed there were (probabilistic) "partial entailments". I'm somewhat skeptical [6] (as are most modern Bayesians).
- Hempel's theory of confirmation [8] satisfies the following:

(SCC) If E confirms H, then E confirms all consequences of H.

- In LFP$_1$, Carnap describes a counterexample to Hempel's (SCC), which presupposes a more ($\star$)-like qualitative conception of confirmation. There, he presupposes:

  Qualitative. E confirms H iff $\Pr(H \mid E) > \Pr(H)$.
- This *probabilistic relevance* conception *violates* (SCC), whereas the previous Pr-threshold conception *implies* (SCC).
- Popper [12] notes this tension in LFP. Largely in response to Popper, Carnap wrote a second edition of LFP [2], which includes a preface acknowledging an "*ambiguity*" in LFP$_1$:
  - **Firmness**. The degree to which E confirms$_f$ H:

    $$\mathfrak{c}_f(H,E) = \Pr(H \mid E).$$
  - **Increase in Firmness**. The degree to which E confirms$_i$ H:

    $$\mathfrak{c}_i(H,E) = \mathfrak{f}[\Pr(H \mid E), \Pr(H)]$$

  $\mathfrak{f}$ measures "the degree to which E *increases* the Pr of H."
- The 1st ed. of LFP was mainly about firmness, and the 2nd edition only adds the preface, which says very little about $\mathfrak{c}_i$. Specifically, no function $\mathfrak{f}$ is rigorously defended there.

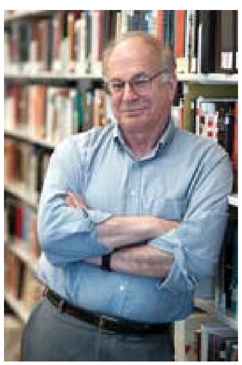
- $\mathfrak{c}_i$ is more similar to ($\ast$) than $\mathfrak{c}_f$ is. To see this, note that we can have $\Pr(H \mid E) > t$ *even if* E ***lowers*** *the probability of* H.
- Example: Let H be the hypothesis that John does *not* have HIV, and let E be a *positive* test result for HIV from a highly reliable test. Plausibly, in such cases, we could have both:
  - $\Pr(H \mid E) > t$, for just about any threshold value t, but
  - $\Pr(H \mid E) < \Pr(H)$, since E *lowers* the probability of H.
- So, if we adopt Carnap's $\mathfrak{c}_f$-explication, then we must say that E confirms H in such cases. But, in ($\ast$)-terms, this implies E provides some *positive evidential support for* H!
- I take it we don't want to say *that*. Intuitively, what we want to say here is that, while H is (still) *highly probable given* E, (nonetheless) E provides (strong!) evidence ***against*** H.
- Carnap [2] seems to appreciate this dissonance, when he concedes $\mathfrak{c}_i$ is (in some settings) "more interesting" than $\mathfrak{c}_f$.
- Contemporary Bayesians would agree with this. They've since embraced a probabilistic relevance conception [13].

Isaac Levi

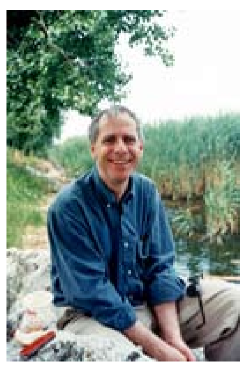
Jim Joyce

- Bayesianism is based on the assumption that the degrees of belief (or credences) of rational agents are *probabilities*.
- Let $\Pr(H)$ be the degree of belief that a rational agent a assigns to H at some time t (call this a's “prior” for H).
- Let $\Pr(H \mid E)$ be the degree of belief that a would assign to H (just after t) were a to learn E at t (a's “posterior” for H).
- Toy Example: Let H be the proposition that a card sampled from some deck is a ♠, and E assert that the card is black.
- Making the standard assumptions about sampling from 52-card decks, $\Pr(H) = \frac{1}{4}$ and $\Pr(H \mid E) = \frac{1}{2}$. So, learning that E *raises the probability* one (rationally) assigns to H.
- Following Popper [12], Bayesians define confirmation in a way that is *formally* very similar to Carnap's $\mathfrak{c}_i$-explication.
- For Bayesians, E confirms H for an agent a at a time t iff $\Pr(H \mid E) > \Pr(H)$, where Pr captures a's credences at t.
- While this is *formally* very similar to Carnap's $\mathfrak{c}_i$, it uses credences as opposed to “logical” probabilities [13], [6].

- When it comes to *quantitative* judgments, Bayesians use various *relevance measures* $\mathfrak{c}$ of degree of confirmation.
- These are much like the candidate functions $\mathfrak{f}$ we saw in connection with Carnapian $\mathfrak{c}_i$, but defined relative to subjective probabilities rather than “logical” probabilities.
- There are *many* ***comparatively*** *distinct* measures. See [5] and [17] for philosophical and psychological discussion.
- Once we choose a measure $\mathfrak{c}(H, E)$ of the degree to which E confirms H, we can explicate **comparative** confirmation relations. *E.g.*, E favors H_1 over H_2 iff $\mathfrak{c}(H_1, E) > \mathfrak{c}(H_2, E)$.
- Note: $\Pr(H \mid E)$ is a *bad* candidate for $\mathfrak{c}(H, E)$ in this context. It implies “E favors H_1 over H_2,” in some cases where E is negatively relevant to H_1 but positively relevant to H_2 [12]!
- In the context of **comparative** confirmation, there is ongoing philosophical/theoretical debate about the appropriate choice of $\mathfrak{c}$ (*e.g.*, the Likelihoodism debate [7]).
- An account is *robust* if it does not depend on choice of $\mathfrak{c}$.



Kahneman



Tversky

- Tversky and Kahneman [19] discuss the following example, which was the first example of the “conjunction fallacy”:
  - (E) Linda is 31, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice and she also participated in antinuclear demonstrations.
- Is it more probable, given E, that Linda is (H_1) a bank teller, or (H_1 and H_2) a bank teller *and* an active feminist?
- Most say “H_1 and H_2” is more probable (given E) than H_1. On its face, this violates comparative probability theory, since $X \vDash Y$ implies $\Pr(X \mid E) \leq \Pr(Y \mid E)$, and $H_1 \,\&\, H_2 \vDash H_1$.
- Experiments have been done to ensure subjects understand “H_1 and H$_2$” in the experiment as a *conjunction* $H_1 \,\&\, H_2$, and H_1 as a *conjunct* thereof (*not* as $H_1 \,\&\, {\sim}H_2$) [15, 16].
- At the same time, the “fallacy” persists when people are queried about *betting odds* rather than *probabilities* [15, 1].
- Comparative Bayesian confirmation can be helpful [11]. We're developing detailed accounts along these lines [4].

- It *is* possible to have $\mathfrak{c}(H_1 \,\&\, H_2, E) > \mathfrak{c}(H_1, E)$ *even though* $H_1 \,\&\, H_2 \vDash H_1$. And, intuitively, this is true in the Linda case.
- As Tversky & Kahneman *themselves* [19] say: "feminist bank teller is a better hypothesis about Linda than bank teller".
- Comparative Bayesian confirmation theory can explain why:

  Theorem. *For all Bayesian relevance measures* $\mathfrak{c}$, *if*
  (i) $\mathfrak{c}(H_2, E \mid H_1) > 0$ *and*
  (ii) $\mathfrak{c}(H_1, E) \leq 0$,
  then $\mathfrak{c}(H_1 \,\&\, H_2, E) > \mathfrak{c}(H_1, E)$.

- Here, $\mathfrak{c}(H_2, E \mid H_1)$ is the degree to which E confirms H_2 (according to $\mathfrak{c}$) *given that the agent already knows* H_1.
- A *logically weaker* pair suffices for $\mathfrak{c}(H_1 \,\&\, H_2, E) > \mathfrak{c}(H_1, E)$. Here is a sharper theorem (based on the (WLL) in [9]):

  Theorem. *For all Bayesian relevance measures* $\mathfrak{c}$, *if*
  (i) $\Pr(E \mid H_1 \,\&\, {\sim}H_2) < \Pr(E \mid H_1 \,\&\, H_2)$ *and*
  (ii*) $\Pr(E \mid H_1 \,\&\, {\sim}H_2) \leq \Pr(E \mid {\sim}H_1)$,
  then $\mathfrak{c}(H_1 \,\&\, H_2, E) > \mathfrak{c}(H_1, E)$.

- The first inequality (i) has already been empirically well established in several traditional (Linda-like) CF cases [14].
- Our (ii)/(ii*) have not been explicitly tested. But, we suspect these will obtain (empirically) in the the traditional CF cases.
- We are performing experiments to test the (i)/(ii) and (i)/(ii*) accounts of the traditional CF cases [4]. Preliminary results indicate that (ii) is commonly endorsed by subjects.
- Interestingly, many seem to judge (i) & (ii) as more plausible than (i) & (ii*) [(ii) *vs* (ii*)]. Do you? Note: (i) & (ii) $\vDash$ (ii*)!
- This suggests (i) & (ii) may provide a more robust explanation than (i) & (ii*) for traditional CFs (a *meta*-CF?).
- But, there are other (non-traditional) sorts of CF cases in which (ii)/(ii*) seem false, and (i) alone does not seem sufficient to predict all patterns of response (next slide).
- Even in this broader class of CFs, however, we think that *some* confirmation-theoretic conditions will be useful for predicting and explaining observed patterns of response.

- Here is a non-traditional CF example: E = John is Scandanavian; H_1 = John has blue eyes; H_2 = John has blond hair.
- In this case, (i) seems plausible, but (ii)/(ii*) do not.
- Moreover, it is not at all clear whether this is (*normatively*) a case in which we *should* have $\mathfrak{c}(H_1 \,\&\, H_2, E) > \mathfrak{c}(H_1, E)$.
- *Descriptively*, we suspect confirmation-theoretic relations between H_1 and H_2 *themselves* may be involved in the CF.
- Specifically, the terms $\mathfrak{c}(H_i, H_j)$ seem to be salient. We bet they are *explanatorily* relevant. There is some preliminary evidence which supports this conjecture [18].
- Psychologically, we think there are two important sets of confirmation-theoretic factors involved in CF cases:
  - $\mathfrak{c}(H_1, E)$, $\mathfrak{c}(H_2, E)$, $\mathfrak{c}(H_1, E \mid H_2)$, $\mathfrak{c}(H_2, E \mid H_1)$. [Traditional CF]
  - $\mathfrak{c}(H_1, H_2)$, $\mathfrak{c}(H_2, H_1)$, $\mathfrak{c}(H_1, H_2 \mid E)$, $\mathfrak{c}(H_2, H_1 \mid E)$. [NT CF]
- More general confirmaiton-theoretic models have recently been developed which seem to subsume and explain all known instances of the "conjunction fallacy" [18].

[1] N. Bonini, K. Tentori and D. Osherson, (2004), "A different conjunction fallacy," *Mind & Lang.*
[2] R. Carnap, (1962), *Logical foundations of probability*, 2nd ed., U. Chicago Press.
[3] ——, (1950), *Logical foundations of probability*, 1st ed., U. Chicago Press.
[4] V. Crupi, B. Fitelson and K. Tentori, (2008), "Probability, Confirmation, and the Conjunction Fallacy," *Thinking and Reasoning.*
[5] B. Fitelson, (2001), *Studies in Bayesian confirmation theory*, PhD. thesis, University of Wisconsin, URL: `http://fitelson.org/thesis.pdf`.
[6] ——, (2005), "Inductive Logic," in *Philosophy of Science: An Encyclopedia.* URL: `http://fitelson.org/il.pdf`.
[7] ——, (2012), "Contrastive Bayesianism," in *Constrastivism in Philosophy*, URL: `http://fitelson.org/cb.pdf`.
[8] C. Hempel, (1945), "Studies in the logic of confirmation," *Mind.*
[9] J. Joyce, (2003), "Bayes' Theorem," *Stanford Encyclopedia of Philosophy.*
[10] J. Keynes, *A treatise on probability*, Macmillan, London, 1921.
[11] I. Levi, (1985), "Illusions about Uncertainty", *British J. for the Philosophy of Science.*
[12] K. Popper, (1954), "Degree of confirmation," *British J. for the Philosophy of Science.*
[13] W. Salmon, (1975), "Confirmation and relevance," in *Induction, Probability, and Confirmation*, Maxwell and Anderson *eds.*, University of Minnesota Press.
[14] E. Shafir, E. Smith and D. Osherson, "Typicality and reasoning fallacies," *Mem. & Cognition.*
[15] A. Sides, D. Osherson, N. Bonini, and R. Viale, (2002), "On the reality of the conjunction fallacy," *Memory and Cognition.*
[16] K. Tentori, N. Bonini, and D. Osherson, (2004), "The conjunction fallacy: a misunderstanding about conjunction?," *Cognitive Science.*
[17] K. Tentori, V. Crupi, N. Bonini and D. Osherson, (2006), "Comparison of confirmation measures," *Cognition.*
[18] K. Tentori, V. Crupi, and S. Russo, (2013), "On the determinants of the conjunction fallacy: Probability *vs.* inductive confirmation," *Journal of Experimental Psychology: General.*
[19] A. Tversky and D. Kahneman, (1983), "Extensional *vs.* intuitive reasoning: the conjunction fallacy in probability judgment," *Psychogical Review.*