Branden Fitelson
Department of Philosophy
Institute for Cognitive and Brain Sciences
&
Group in Logic and the Methodology of Science
University of California–Berkeley
branden@fitelson.org
http://fitelson.org/

Carnap [1] aims to provide a formal explication of an informal concept (relation) he calls “confirmation”.

He clarifies “$E$ confirms $H$” in various ways, including:

($\ast$) $E$ provides some positive evidential support for $H$.

His formal explication of “$E$ confirms $H$” (in [1]) is:

(1) $E$ confirms $H$ iff $Pr(H \mid E) > r$, where $Pr$ is a suitable (“logical”) probability function, and $r$ is a threshold value.

Unfortunately, Carnap [1] is not entirely consistent in his formal analyses and applications of confirmation.

Popper [11] points out that in some parts of [1], Carnap has a different explication of confirmation in mind, namely:

(2) $E$ confirms $H$ iff $Pr(H \mid E) > Pr(H)$, where $Pr$ is a suitable (“logical”) probability function. [i.e., correlation under Pr]

In response to Popper, Carnap [2] postulated an ambiguity in the concept of confirmation [(1)- vs (2)-confirmation].

To some modern readers (e.g., me), this seems inadequate, since (2) seems to be a better explication of the informal concept ($\ast$) that Carnap aimed to explicate in the first place.

To see why (2) is more similar to ($\ast$) than (1) is, note that (1) can be satisfied 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:

- $Pr(H \mid E) > r$, for just about any threshold value $r$, but
- $Pr(H \mid E) < Pr(H)$, since $E$ lowers the probability of $H$.

So, if we adopt Carnap’s (1)-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 $H$.

Rather than ambiguity, I’d say this reflects confusion about the nature of the concept ($\ast$) Carnap was trying to explicate.

Even Carnap [2] says (2) is “more interesting” than (1).

Contemporary (Bayesian) confirmation theorists seem to agree. They no longer think of confirmation in (1)-terms …
Bayesianism assumes that the *epistemic* degrees of belief (that is, the *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) $E$ (or supposing that $E$) *raises the probability of* $H$.

- Following Popper [11], Bayesians define confirmation in a way that is *formally* very similar to Carnap’s (2)-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 (2), it does not assume that there are objective, “logical” probabilities.

---

**Fact.** No two of $\{d, r, l, s\}$ are ordinally equivalent.

- OK, but do they disagree on *important* applications or in *important* cases? Unfortunately, they disagree *radically*.

---

**Fact.** *Almost every* argument/application in the literature is valid for *only some* choices of $d, r, l, s$. I have called this the *problem of measure sensitivity*. See my [4] for a survey.

- We need some normative principles to narrow the field …

---

There are *many logically equivalent* (but syntactically distinct) ways of saying $E$ confirms $H$, in the Bayesian sense.

- Here are the three most common ways:
  - $E$ confirms $H$ iff $\Pr(H \mid E) > \Pr(H)$. [2] $> \frac{1}{2}$
  - $E$ confirms $H$ iff $\Pr(E \mid H) > \Pr(E \mid \neg H)$. [1] $> \frac{1}{2}$
  - $E$ confirms $H$ iff $\Pr(H \mid E) > \Pr(H \mid \neg E)$. [$\frac{1}{2} > 0$]

- By taking differences or ratios of the left/right sides of such inequalities, various confirmation *measures* $c(H, E)$ emerge.

- A plethora of such confirmation measures have been used in the literature of Bayesian confirmation theory. See my thesis [4] for a survey. Here are the four most popular $c$’s:

\[
d(H, E) \equiv \Pr(H \mid E) - \Pr(H)
\]

\[
r(H, E) \equiv \log \frac{\Pr(H \mid E)}{\Pr(H)} = \Pr(H \mid E) - \Pr(H)
\]

\[
l(H, E) \equiv \log \frac{\Pr(E \mid H)}{\Pr(E \mid \neg H)} = \Pr(E \mid H) - \Pr(E \mid \neg H)
\]

\[
s(H, E) \equiv \Pr(H \mid E) - \Pr(H \mid \neg E)
\]

---

**Consider the following two propositions concerning a card $c$, drawn at random from a standard deck of playing cards:**

$E$: $c$ is the ace of spades. $H$: $c$ is some spade.

**I take it as intuitively clear and uncontroversial that:**

- The degree to which $E$ confirms $H$ is the degree to which $H$ confirms $E$, since $E \equiv H$, but $H \not\equiv E$. [$c(H, E) \equiv c(E, H)$]

- The degree to which $E$ confirms $H$ is the degree to which $\neg E$ disconfirms $H$, since $E \equiv H$, $\neg E \equiv \neg H$. [$c(H, E) \equiv -c(H, \neg E)$]

**Therefore, no adequate measure of confirmation $c$ should be such that either $c(H, E) = c(E, H)$ or $c(H, E) = -c(H, \neg E)$ for all $E$ and $H$ and for all probability functions $\Pr$. I’ll call these two symmetry desiderata $S_1$ and $S_2$, respectively.**

**Note:** for all $H$, $E$, and for all $Pr$, $r(H, E) = r(E, H)$ and $s(H, E) = -s(H, \neg E)$. That is, $r$ violates $S_1$ and $s$ violates $S_2$.

- Both $d$ and $l$ satisfy these $S$-desiderata. This narrows the field to $d$ and $l$ [3]. We can narrow the field further still …
If we think of inductive logic as a quantitative generalization of deductive logic, then the following logical desideratum seems natural (it’s also implicit in the previous example):

(†) **Quantitative Rendition.** \( c(H, E) \) should be maximal when \( E \models H \) and \( c(H, E) \) should be minimal when \( E \not\models \sim H \).

(†) **Comparative Rendition.** If \( E = H \) but \( E' \not\models H' \), then the following inequality should hold: \( c(H, E) \geq c(H', E') \).

The measure \( d \) violates these desiderata. For, when \( E = H \):

\[
d(H, E) = \Pr(H | E) - \Pr(H) = 1 - \Pr(H) = \Pr(\sim H)
\]

So, if the prior probability of \( H \) is sufficiently high, then (according to \( d \) \( E \) will confirm \( H \) very weakly, even if \( E \models H \).

From an inductive-logical point of view, this is absurd, since the logical strength of a valid argument should not depend on how probable its conclusion is (or on its truth-value).

Indeed, of all the Bayesian measures of confirmation that have been used in the literature, only \( l \) (or its ordinal equivalents) satisfy all three of our desiderata: \( S_1, S_2, (†) \).

A second example from K&T that’s worth thinking about in this connection is the so-called “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 \( (H_2) \) a bank teller and active in the feminist movement?

Most people answer that \( H_2 \) is more probable (given \( E \)) than \( H_1 \) is. This answer violates Pr-theory, since \( H_2 \models H_1 \).

Note: it is possible to have \( l(H_2, E) > l(H_1, E) \) even if \( H_2 \models H_1 \). And, \( E \) could constitute better evidence for \( H_2 \) than for \( H_1 \). So, again, maybe this poor probability judgment also reflects a good underlying confirmation judgment [12].

I do not mean to say that K&T’s normative assessments about Pr are wrong. But, I do want to suggest that people may be better at making confirmation judgments than probability judgments. If we only had some evidence …

Kahneman and Tversky [8] amassed lots of data, which they claimed indicated violations of normative principles for probability judgments (i.e., violations of the Pr-axioms).

If Carnap was confused (along with many others) about the probability/confirmation distinction, could this confusion also underlie some of these erroneous Pr judgments?

Two examples from K&T come to mind. First, their experiments on the neglect of “base rate” information.

When people are asked to assess the probability that John has AIDS, given that he tested positive for AIDS according to a very reliable test protocol, they often report high values.

This seems to violate Bayes’s Theorem, since AIDS has such a low base rate (prior?) in the population (and they know this). This does seem to be a poor probability judgment [9].

But, could this also reflect a good underlying confirmation or evidential support judgment? Note: \( l(H, E) \) is very close to the value reported by experts in these examples [6].

A recent study [10] was designed to answer this question …

Amazingly, until very recently there have been almost no psychological studies on how people actually make confirmation judgments (in the present, Bayesian sense).

This was surprising to me, mainly for the following reasons:

- Because of the long-standing confusion about probability vs confirmation in the philosophical literature, I thought that this should be a ripe area for psychological research.
- I’ve suspected that confirmation judgments should be more robust than Pr-judgments, since they are (normatively) less sensitive to subjective factors (in particular, “priors” [5]).
- I am happy to report that this now seems to be evolving into a ripe area for psychological research. Dan Osherson and his colleagues are largely responsible for this change.
- One thing we’d like to know is whether people tend to make quantitative judgments of confirmational strength that accord with normatively adequate measures like \( l \).
- A recent study [10] was designed to answer this question …
As far as I know, the forthcoming study by Osherson et al. [10] is the first designed explicitly to test Bayesian measures of confirmation against each other for descriptive accuracy. Their study involved 24 undergraduates (U. of Trento). They were (individually) faced with the following scenario.

- They were shown two opaque urns (A, B), where A contains 30/10 black/white balls, and B contains 15/25 B/W balls.
- A fair coin was tossed, and an urn selected at random. Then, 10 balls were drawn (at random) without replacement.
- After each draw, they were asked to rank the evidential impact of that draw on the hypotheses (a) that A was chosen, and (b) that B was chosen, on a scale with 7 “ticks”.
- Tick 1: “weakens my conviction extremely”, tick 7: “strengthens my conviction extremely”. Tick 4: “no effect”.
- Then, the subject was asked to estimate probabilities $\Pr(A \mid E)$ and $\Pr(B \mid E)$ and likelihoods $\Pr(E \mid A)$ and $\Pr(E \mid B)$.
- Finally, these subjective estimates of probabilities and likelihoods were plugged-in to the various measures of confirmation. And, correlation statistics were calculated.

First, I would suggest looking at comparative/relation confirmation judgments, rather than quantitative ones. I suspect these will be even more robust and objective [5].

Second, I would suggest controlling for certain other pragmatic factors that may confound (or create) differences between measures. Jim Joyce has discussed such factors [7].

Third, the protocol of Osherson et al. was unable to test the descriptive accuracy of the measure $s$. It would be nice to generalize their protocol to include $s$ (and others like it).

Finally, I would also like to see some experiments designed explicitly to distinguish qualitative confirmation judgments from probability-threshold judgments [Carnapian (1) vs (2)].

I suspect that people’s judgments about “what confirms what” come apart sharply from their judgments of what is “probable”. But, it would be nice to have more data on this.

E.g.: I bet jurors who learn their (guilty) verdict was false will retract “probable” claims, not “supported-by-E” claims.

The experimenters also plugged-in objective probabilities and likelihoods, to see what predictions those yielded.

The results were (to me) somewhat (pleasantly!) surprising:

- Of the measures $d$, $r$, and $l$, the measure $l$ was significantly better at predicting confirmation judgments, both using the subjective and the objective probabilities and likelihoods.
- Note: their protocol was unable to test the accuracy of $s$.
- Several additional measures from the literature were tested, and $l$ was significantly better than all of the other measures, when objective probabilities/likelihoods were used.
- $l$ was not significantly worse than any other measure tested, when subjective probabilities/likelihoods were used.
- The posterior probabilities (either objective or subjective) were very poor predictors. This indicates that the subjects distinguished confirmation & probability [Carnap’s (1) & (2)].

This (plus subj ≠ obj) confirms what I have long suspected: people are better at making confirmation judgments than probability judgments. Of course, more studies are needed.

Now, for some research suggestions from the armchair . . .