Discussion: Re-solving Irrelevant Conjunction with Probabilistic Independence*

James Hawthorne and Branden Fitelson†

Naive deductivist accounts of confirmation have the undesirable consequence that if E confirms H, then E also confirms the conjunction $H \cdot X$, for any X—even if X is completely irrelevant to E and H. Bayesian accounts of confirmation may appear to have the same problem. In a recent article in this journal Fitelson (2002) argued that existing Bayesian attempts to resolve of this problem are inadequate in several important respects. Fitelson then proposes a new-and-improved Bayesian account that overcomes the *problem of irrelevant conjunction*, and does so in a more general setting than past attempts. We will show how to simplify and improve upon Fitelson's solution.

1. Introduction. We will begin by recalling the *problem of irrelevant conjunction*. Then we describe Fitelson's (2002) solution. And finally we show how to improve on it.

The problem of irrelevant conjunction was originally raised as a problem for hypothetico-deductive (H-D) accounts of confirmation. On H-D accounts, E confirms H relative to background K when $H \cdot K$ deductively entails E (i.e., if $H \cdot K \vDash E$). This leads to the following result:

(1) If E H-D-confirms H relative to K, then E H-D-confirms $H \cdot X$ relative to K, for any X.

The problem with (1) is that when E confirms H, any other hypothesis X compatible with H gets a free confirmational ride. Merely tack X onto H, and E confirms them together, regardless of the fact that X may be *utterly irrelevant* to E in the presence of H.

505

Bayesian confirmation may seem to have an advantage over the H-D account regarding irrelevant conjunctions. For, the Bayesian account does not generally have the property expressed in (1). On a Bayesian account, "E confirms H relative to K" just means that "$\Pr(H|E \cdot K) > \Pr(H|K)$" for an appropriate Bayesian probability function Pr. And in lots of cases evidence E may confirm H in this probabilistic way without confirming some conjunction $H \cdot X$—i.e. without also making $\Pr(H \cdot X|E \cdot K) > \Pr(H \cdot X|K)$. So, Bayesianism is generally immune to the original problem of irrelevant conjunction.

However, Bayesian confirmation still seems to suffer from this problem in the case of *deductive evidence*. That is, Bayesian confirmation and H-D confirmation both satisfy the following *special case* of (1) whenever E is less than certain on K—i.e. when $\Pr(E|K) < 1$:

(1′) If $H \cdot K \vDash E$, then E confirms $H \cdot X$ relative to K for *any* X consistent with $H \cdot K$.

That is, when $H \cdot K \vDash E$ we have, for any X, $\Pr(E|H \cdot X \cdot K) = 1$. So, provided X is logically consistent with $H \cdot K$ and is probabilistically consistent with it as well (i.e. $\Pr(H \cdot X \cdot K) > 0$), we have

$$\Pr(H \cdot X|E \cdot K) = \Pr(E|H \cdot X \cdot K)\Pr(H \cdot X|K)/\Pr(E|K)$$

$$> \Pr(H \cdot X|K).$$

So, Bayesians must concede that the problem persists for evidence that is deductively entailed by $H \cdot K$.

Bayesians do have some wiggle room, however. They can concede (1′), but argue that in the context of deductive evidence, H simpliciter will always be better confirmed than $H \cdot X$. One way to attempt to legitimate this claim is to point out that it follows from the axioms of probability that $\Pr(H|E \cdot K) \geq \Pr(H \cdot X|E \cdot K)$, and that when $H \cdot K \vDash E$ (and $\Pr(H \cdot X \cdot K) > 0$), equality only holds in the special case where $\Pr(X|H \cdot K) = 1$ (since

$$\Pr(H \cdot X|E \cdot K) = \Pr(X|H \cdot E \cdot K)\Pr(H|E \cdot K)$$

$$= \Pr(X|H \cdot K)\Pr(H|E \cdot K)).$$

So, in the deductive case H is always made at least as probable as $H \cdot X$ by E (relative to K), and in the interesting cases (where $\Pr(X|H \cdot K) \neq 1$), it's made more probable.

This solution turns on reading the "E confirms H_1 more than H_2" relation as "H_1 is more probable given E than is H_2 given E." But that is not generally a good way to understand "confirms more." For, given this reading, in more general cases it will often happen that "E confirms H_2 more than H_1 relative to K" (because $\Pr(H_2|E \cdot K) > \Pr(H_1|E \cdot K)$),

while E actually lowers the probability of H_2 ($\Pr(H_2|E \cdot K) < \Pr(H_2|K)$) and raises the probability of H_1 ($\Pr(H_1|E \cdot K) > \Pr(H_1|K)$). So, the brute comparison of the relative sizes of posterior probabilities is an intuitively unappealing way to cash out the "confirms more" relation. Thus, the proposed solution to *irrelevant conjunction* depends on a flawed account of "confirms more."

So, is there a better way to cash out "confirms more" that gives the desired solution to *irrelevant conjunction*? Several accounts of "*E* confirms *H* given *K*" have been proposed in the literature that provide measures of how much *E incrementally increases* the probability of *H*, given *K*—increases it over the probability of *H* prior to the evidence. The most common are the ratio measure, the difference measure, and the likelihood ratio measure:

ratio measure:

$$r(H,\,E|K) = \frac{\Pr(H|E \cdot K)}{\Pr(H|K)} = \frac{\Pr(E|H \cdot K)}{\Pr(E|K)};$$

difference measure:

$$d(H,\,E|K) = \Pr(H|E \cdot K) - \Pr(H|K);$$

likelihood-ratio measure:

$$l(H,\,E|K) = \frac{\Pr(E|H \cdot K)}{\Pr(E|{\sim}H \cdot K)}.$$

It should be clear how these measures might be applied to the problem of irrelevant conjunction. The ratio measure would have it that *E* confirms H_2 better than H_1 (relative to *K*) just when the ratio $r(H_2,\,E|K)$ is larger than the ratio $r(H_1,\,E|K)$. The difference measure says that *E* confirms H_2 better than H_1 (relative to *K*) just when the difference $d(H_2,\,E|K)$ is larger than the difference $d(H_1,\,E|K)$. And likelihood-ratio measure says that *E* confirms H_2 better than H_1 (relative to *K*) just when $l(H_2,\,E|K)$ is larger than $l(H_1,\,E|K)$. Thus, a Bayesian solution to the problem of irrelevant conjunction might be obtained if it can be shown that $c(H,\,E|K)$ is larger than $c(H \cdot X,\,E|K)$ where *c* is some appropriate measure of incremental confirmation such as *r*, *d*, or *l*.

It is now well known that when each of these measures is applied to the same issue or problem in Bayesian treatments of confirmation, they may provide divergent results. Indeed, Fitelson (1999, 2001) has extensively analyzed this issue, and finds strong grounds to prefer *l* over other measures. Be that as it may, what we want to know is whether the irrelevant conjunction problem has a Bayesian solution based on any, or all of these measures. The Fitelson paper under discussion provides an answer

to this query. Indeed, Fitelson's answer goes beyond the call. For, Fitelson holds, a proper Bayesian solution to the problem of irrelevant conjunction should be directly extendable to the case of non-deductive evidence as well. And Fitelson's analysis applies to non-deductive cases.

2. Fitelson's Original Solution. Fitelson suggests that a proper Bayesian analysis of the problem of irrelevant conjunction should start by saying what, in a Bayesian framework, it means for X to be irrelevant to the confirmation of H by E, given K. He suggests that this should mean that X is irrelevant to H, E, and logical combinations of H and E, where *irrelevance* is to be understood in the usual Bayesian sense of probabilistic independence. Formally, the proposal is this:

> **Definition. Confirmational Irrelevance**. A is *confirmationally irrelevant* to B relative to K just when A and B are probabilistically independent, given K i.e., when $Pr(A \cdot B|K) = Pr(A|K) \cdot Pr(B|K)$.

Fitelson then proves two theorems that bear on the problem of irrelevant conjunction for both deductive and non-deductive evidence. The first theorem shows that the apparent problem persists even in non-deductive cases—that if E confirms H, but X is confirmationally irrelevant to H, E, and $H \cdot E$, then on each of the measures r, d, and l, E also confirms (i.e. incrementally raises the probability of) $H \cdot X$.

> **Theorem 1**. If E confirms H, and X is confirmationally irrelevant to H, E, and $H \cdot E$ (relative to background K), then E also confirms $H \cdot X$ (relative to K).

Fitelson's second theorem then comes to the rescue of the Bayesian account by showing that if E confirms H, but X is confirmationally irrelevant to H, E, and $H \cdot E$, then on measures d and l (but not on r) E incrementally raises the probability of H *more* than it does $H \cdot X$.

> **Theorem 2**. If E confirms H, and X is confirmationally irrelevant to H, E, and $H \cdot E$ (relative to background K) and $Pr(X|K) \neq 1$, then $c(H, E|K) > c(H \cdot X, E|K)$, where c may be either the difference measure d or the likelihood-ratio measure l of degree of confirmation (but, c may *not* be the ratio measure r, since in cases of irrelevant conjunction we will have $r(H, E|K) = r(H \cdot X, E|K)$).

This result seems quite satisfactory, but for one thing. The antecedents of these theorems appear to be a little too strong, and perhaps less intuitively compelling than we might like.

3. An Improvement on the Solution. In the deductive case the intuition about the irrelevance of X flows from the idea that a hypothesis is tested

by what it *says* about the evidence. Adding X to H (given K) says *nothing more* about E than H (given K) already says. So, intuitively, E should confirm $H \cdot X$ (given K) no more than it confirms H alone (given K). Indeed, intuitively E should perhaps confirm $H \cdot X$ less than H alone (given K). A natural Bayesian extension of this idea about X adding nothing to what H says about E (given K) is this:

> **Natural Conjunctive Irrelevance Criterion**. X is an irrelevant conjunct to H given K with respect to evidence E just in case $Pr(E|H \cdot X \cdot K) = Pr(E|H \cdot K)$ i.e. just when E is independent of X given H together with K.

This criterion involves an intuition about probabilistic independence that is central to Fitelson's previous theorems, but draws on much more intuitively direct version of the idea that X is an irrelevant conjunct. This describes one sense in which the antecedents in Fitelson's theorems seem overly strong.

The antecedents of the theorems are too strong in a more literal sense as well. For it turns out that:

i. If "X is confirmationally irrelevant to H, E, and $H \cdot E$" in the sense employed in the theorems, then "X is an irrelevant conjunct to H given K with respect to evidence E" in the sense of the Natural Conjunctive Irrelevance Criterion just defined. And the Natural Conjunctive Irrelevance Criterion is strictly weaker than supposing X to be confirmationally irrelevant to H, E, and $H \cdot E$. Furthermore:

ii. The Natural Conjunctive Irrelevance Criterion suffices in place of Fitelson's original irrelevance conditions, to establish of the consequents of the theorems.

To see that claim (i) holds, notice that Fitelson's original independence conditions imply that

$$\Pr(E|H \cdot X \cdot K) \cdot \Pr(H|K) \cdot \Pr(X|K)$$
$$= \Pr(E|H \cdot X \cdot K) \cdot \Pr(H \cdot X|K)$$
$$= \Pr(H \cdot X \cdot E|K)$$
$$= \Pr(H \cdot E|K) \cdot \Pr(X|K)$$
$$= \Pr(E|H \cdot K) \cdot \Pr(H|K) \cdot \Pr(X|K);$$

so $\Pr(E|H \cdot X \cdot K) = \Pr(E|H \cdot K)$. But Fitelson's original condition is not equivalent to the revised condition, since we can easily have cases where $\Pr(E|H \cdot X \cdot K) = \Pr(E|H \cdot K)$ while $\Pr(X|H \cdot K) \neq \Pr(X|K)$ (i.e. $\Pr(H \cdot X|K) \neq \Pr(X|K)\Pr(H|K)$). Consider, for example, cases where X

is a kind of evidence statement E^* (e.g. "the second toss will be *heads*") and hypothesis H is relevant to E^* (e.g. H says "the coin is fair" and $Pr(E^*|H \cdot K) \neq P(E^*|K)$), and where E (e.g. "the first toss is *heads*") is independent of E^* given $H \cdot K$.

To verify claim (ii), first observe that the following revised version of Theorem 1 holds:

> **Revised Theorem 1**. If E confirms H relative to K and $Pr(E|H \cdot X \cdot K) = Pr(E|H \cdot K)$, then E also confirms $H \cdot X$ relative to K, where the notion of *confirmation* is given by either the ratio measure r or the difference measure d or the likelihood-ratio measure l.

This is a simplified version of the Bayesian "bad news" theorem. In the case of the ratio measure r it says that if $Pr(E|H \cdot X \cdot K) = Pr(E|H \cdot K)$ and $r(H, E|K) > 1$ (i.e. H is confirmed by E given K on measure r), then $r(H \cdot X, E|K) > 1$ (i.e. $H \cdot X$ is confirmed by E given K on measure r). For the difference measure d the revised theorem says that if $Pr(E|H \cdot X \cdot K) = Pr(E|H \cdot K)$ and $d(H, E|K) > 0$ (i.e. H is confirmed by E given K on measure d), then $d(H \cdot X, E|K) > 0$ (i.e. $H \cdot X$ is confirmed by E given K on measure d). In the case of the likelihood-ratio measure l the revised theorem says that if $Pr(E|H \cdot X \cdot K) = Pr(E|H \cdot K)$ and $l(H, E|K) > 1$ (i.e. H is confirmed by E given K on measure l), then $l(H \cdot X, E|K) > 1$ (i.e. $H \cdot X$ is confirmed by E given K on measure l).

In spite of the bad news from Revised Theorem 1, the corresponding revised version of Theorem 2 shows that even on the weaker, more intuitive notion of irrelevance, two of the three Bayesian measures of incremental confirmation say that hypotheses are always better confirmed than their conjunctions with irrelevant alternatives.

> **Revised Theorem 2**. If E confirms H relative to K and $Pr(E|X \cdot H \cdot K) = Pr(E|H \cdot K)$ and $Pr(X|H \cdot K) \neq 1$, then $c(H, E|K) > c(H \cdot X, E|K)$, where c may be either the difference measure d or the likelihood-ratio measure l of degree of incremental confirmation. But, c may *not* be the ratio measure r, since for r, when E confirms H relative to K and $Pr(E|X \cdot H \cdot K) = Pr(E|H \cdot K)$, we have that $r(H, E|K) = r(H \cdot X, E|K)$.

This revision of Fitelson's original Bayesian resolution of the irrelevant conjunction problem significantly strengthens that result. Like Fitelson's earlier version, this resolution is not restricted to the special case of deductive evidence. However, this version has the advantage of employing a notion of an irrelevant conjunct that is precisely analogous to that in the original deductive evidence case. It subsumes the deductive irrelevant conjunction issue as a special case in a natural way. And, like Fitelson's

original approach, this solution is more robust, more measure-insensitive, than other suggested resolutions of the problem.

Appendix: Proofs of Theorems

Proof of Revised Theorem 1

Suppose that $\Pr(E|X \cdot H \cdot K) = \Pr(E|H \cdot K)$ in each case, 1, 2, and 3, below.

1. For ratio measure r: If $r(H, E|K) > 1$ (i.e. H is confirmed by E given K on measure r), then

$$1 < r(H, E|K) = \frac{\Pr(H|E \cdot K)}{\Pr(H|K)} = \frac{\Pr(E|H \cdot K)}{\Pr(E|K)}$$

$$= \frac{\Pr(E|H \cdot X \cdot K)}{\Pr(E|K)} = r(H \cdot X, E|K)$$

(i.e. $H \cdot X$ is confirmed by E given K on measure r).

2. For the difference measure d: Suppose $d(H, E|K) > 0$ (i.e. H is confirmed by E given K on measure d). Then $0 < d(H, E|K) = \Pr(H|E \cdot K) - \Pr(H|K)$, so

$$1 < \frac{\Pr(H|E \cdot K)}{\Pr(H|K)} = \frac{\Pr(E|H \cdot K)}{\Pr(E|K)}$$

$$= \frac{\Pr(E|H \cdot X \cdot K)}{\Pr(E|K)} = \frac{\Pr(H \cdot X|E \cdot K)}{\Pr(H|K)},$$

so $0 < \Pr(H \cdot X|E \cdot K) - \Pr(H \cdot X|K)$ (i.e. $H \cdot X$ is confirmed by E given K on measure d).

3. For the likelihood-ratio measure l: Suppose $l(H, E|K) > 1$ (i.e. H is confirmed by E given K on measure l). Then

$$1 < l(H, E|K) = \frac{\Pr(E|H \cdot K)}{\Pr(E|\sim H \cdot K)};$$

so $\Pr(E|\sim H \cdot K) < \Pr(E|H \cdot K)$. This implies $1 < \Pr(E|H \cdot K)/\Pr(E|K)$ [because

$$\Pr(E|K) =$$

$$\Pr(E|H \cdot K) \cdot (1 - \Pr(\sim H|K)) + \Pr(E|\sim H \cdot K \cdot \Pr(\sim H|K),$$

so

$$\Pr(E|K) - \Pr(E|H \cdot K) =$$

$$(-\Pr(E|H \cdot K) + \Pr(E|{\sim}H \cdot K)) \cdot \Pr({\sim}H|K) < 0].$$

Then,

$$1 < \frac{\Pr(E|H \cdot X \cdot K)}{\Pr(E|K)} = \frac{\Pr(H \cdot X|E \cdot K)}{\Pr(H \cdot X|K)}.$$

So we have $\Pr(H \cdot X|K) < \Pr(H \cdot X|E \cdot K)$, which implies

$$\Pr({\sim}(H \cdot X)|K) > \Pr({\sim}(H \cdot X)|E \cdot K).$$

These two inequalities together yield

$$\frac{\Pr(H \cdot X|K)}{\Pr({\sim}(H \cdot X)|K)} < \frac{\Pr(H \cdot X|E \cdot K)}{\Pr({\sim}(H \cdot X)|E \cdot K)}$$

$$= \left[\frac{\Pr(E|H \cdot X \cdot K)}{\Pr(E|{\sim}(H \cdot X) \cdot K)}\right] \cdot \left[\frac{\Pr(H \cdot X|K)}{\Pr({\sim}(H \cdot X)|K)}\right].$$

Thus $1 < \Pr(E|H \cdot X \cdot K)/\Pr(E|{\sim}(H \cdot X) \cdot K)$ (i.e. $H \cdot X$ is confirmed by E given K on measure l).

Proof of Revised Theorem 2

Suppose that $\Pr(E|X \cdot H \cdot K) = \Pr(E|H \cdot K)$ and $\Pr(X|H \cdot K) \neq 1$ in each case, 1, 2, and 3, below.

1. For the $c = r$ case:

$$r(H, E|K) = \frac{\Pr(E|H \cdot K)}{\Pr(E|K)} = \frac{\Pr(H|E \cdot K)}{\Pr(H|K)}$$

$$= \frac{\Pr(E|H \cdot X \cdot K)}{\Pr(E|K)} = r(H \cdot X, E|K).$$

2. For the $c = d$ case: Suppose E confirms H relative to K. Then

$d(H, E|K) > 0$—i.e. $\Pr(H|E \cdot K) - \Pr(H|K) > 0$. Then

$$\Pr(H \cdot X|E \cdot K) = \frac{\Pr(E|H \cdot X \cdot K)\Pr(H \cdot X|K)}{\Pr(E|K)}$$

$$= \left[\frac{\Pr(E|H \cdot K)}{\Pr(E|K)}\right] \cdot \Pr(H \cdot X|K)$$

$$= \left[\frac{\Pr(H|E \cdot K)}{\Pr(H|K)}\right] \cdot \Pr(H \cdot X|K).$$

So,

$$d(H \cdot X, E|K) = \Pr(H \cdot X|E \cdot K) - \Pr(H \cdot X|K)$$

$$= \Pr(H \cdot X|K) \cdot \left[\left[\frac{\Pr(H|E \cdot K)}{\Pr(H|K)}\right] - 1\right]$$

$$= \left[\frac{\Pr(H \cdot X|K)}{\Pr(H|K)}\right] \cdot d(H, E|K)$$

$$= \Pr(X|H \cdot K) \cdot d(H, E|K) < d(H, E|K)$$

(unless $\Pr(X|H \cdot K) = 1$, in which case they're equal).

3. For the $c = l$ case: Suppose E confirms H relative to K. Then $l(H, E|K) > 1$—i.e. $\Pr(E|H \cdot K)/\Pr(E|{\sim}H \cdot K) > 1$. Notice that

$$\Pr(E|{\sim}X \cdot H \cdot K) = \Pr(E|H \cdot K)$$

(since $\Pr(X|H \cdot K) \neq 1$ and $\Pr(E|X \cdot H \cdot K) = \Pr(E|H \cdot K)$). Also notice that $\Pr({\sim}H|{\sim}(X \cdot H) \cdot K) < 1$ (otherwise $0 = \Pr(H|{\sim}(X \cdot H) \cdot K)$, so

$$0 = \Pr(H \cdot {\sim}(X \cdot H)|K) = \Pr(H \cdot {\sim}X|K)$$

$$= \Pr({\sim}X|H \cdot K) \cdot \Pr(H|K),$$

so $\Pr(X|H \cdot K) = 1 - \Pr({\sim}X|H \cdot K) = 1$, which contradicts our assumption that $\Pr(X|H \cdot K) < 1$).

Now

$$\Pr(E|{\sim}(X \cdot H) \cdot K)$$

$$= \Pr(E|H \cdot {\sim}(X \cdot H) \cdot K) \cdot \Pr(H|{\sim}(X \cdot H) \cdot K)$$

$$+ \Pr(E|{\sim}H \cdot {\sim}(X \cdot H) \cdot K) \cdot \Pr({\sim}H|{\sim}(X \cdot H) \cdot K)$$

$$= \Pr(E|H \cdot K) \cdot \Pr(H|{\sim}(X \cdot H) \cdot K)$$

$$+ \Pr(E|{\sim}H \cdot K) \cdot \Pr({\sim}H|{\sim}(X \cdot H) \cdot K).$$

So,

$$\frac{\Pr(E|\sim(X \cdot H) \cdot K)}{\Pr(E|(X \cdot H) \cdot K)} = \Pr(H|\sim(X \cdot H) \cdot K) + \Pr(\sim H|\sim(X \cdot H) \cdot K)$$

$$\cdot \left[\frac{\Pr(E|\sim H \cdot K)}{\Pr(E|H \cdot K)}\right] > \frac{\Pr(E|\sim H \cdot K)}{\Pr(E|H \cdot K)}.$$

REFERENCES

Fitelson, Branden (1999), "The Plurality of Bayesian Measures of Confirmation and the Problem of Measure Sensitivity", *Philosophy of Science* 66 (Proceedings): S362–S378.
——— (2001), *Studies in Bayesian Confirmation Theory*. Ph.D. Dissertation. Madison, WI: University of Wisconsin—Madison.
——— (2002), "Putting the Irrelevance Back Into the Problem of Irrelevant Conjunction", *Philosophy of Science* 69 (4): 611–622.