

## Comments on “On an Unsuccessful Attempt at Filtering Out the One True Measure of Confirmation”

This is a very interesting and well-written paper on an important class of issues arising in recent work on Bayesian confirmation theory. As it stands, however, it is not suitable for publication. But, with sufficient (albeit, non-trivial) work, a future version of it *may* be suitable. Below, I provide detailed comments on the paper with several suggestions for requisite amendments. I would be happy to referee a revised version of this paper, should the author decide they can meet the challenges I set out in these remarks. I think the challenges are quite difficult, but I do hold out some hope that the author can meet them. My remarks will progress (roughly) in typographical order.

1. On page 1, when the author reports the relevance criterion of confirmation, they give a version of it that “leaves out background knowledge” “for simplicity”. This omission is important, as several desiderata in the historical literature are sensitive to the inclusion of background knowledge in the relevance criterion (and in the quantitative measures of relevance confirmation the author discusses). I will return to this point, below.
2. On page 2, the author saddles Fitelson with the claim that the Kemeny-Oppenheim (*KO*) measure (or its ordinal equivalents) is “the one true measure of confirmation”, where this is interpreted as a strong uniqueness claim. In fact, all Fitelson has claimed (in print up to now) is that these *KO*-measures are the only *historically proposed* measures that satisfy the deductive evidence desiderata discussed by the author. Fitelson is well aware that there are other relevance measures satisfying this deductive desideratum. As I will explain, below, when one properly combines the deductive evidence desideratum with the other desiderata in Fitelson’s work (including some that make use of background knowledge, omitted from the author’s discussion), one is able to rule-out not only all the measures that were known to Fitelson before, but also the new and interesting measure proposed by the author in the present paper. I think this fact undermines almost all of the power of the author’s critique of Fitelson.
3. Also on page 2, the author claims that the relevance criterion of confirmation “suggests equating degree of confirmation with *size of probability increase*” rather than *inferential strength*. This seems to suggest that Fitelson may be misinterpreting the relevance criterion itself, since he takes the aim of confirmation theory to be the explication of a relevance measure of inferential strength (as embodied in his deductive evidence desideratum). This charge is unfair to Fitelson and also misleading. Carnap himself was confused on this point, so this is certainly an excusable error. It is an error because there are *many logically equivalent* ways of expressing the relevance criterion (assuming Komogorov’s theory of Pr). Some of these ways may suggest a “*size of probability increase*” reading, but others do not. The author expresses the relevance criterion ( $\mathcal{R}$ ) as:

$\mathcal{R}_1$ :  $E$  confirms  $H$  iff  $\Pr(H | E) > \Pr(H)$ ;  $E$  disconfirms  $H$  iff  $\Pr(H | E) < \Pr(H)$ ; and  $E$  is confirmationally irrelevant to  $H$  iff  $\Pr(H | E) = \Pr(H)$ .

This way of writing down  $\mathcal{R}$  makes it seem as if there is some special significance to the comparison of  $\Pr(H|E)$  and  $\Pr(H)$ . But, we could just as easily write down  $\mathcal{R}$  in the following logically equivalent<sup>1</sup> way:

$\mathcal{R}_2$ :  $E$  confirms  $H$  iff  $\Pr(E|H) > \Pr(E|\neg H)$ ;  $E$  disconfirms  $H$  iff  $\Pr(E|H) < \Pr(E|\neg H)$ ; and  $E$  is irrelevant to  $H$  iff  $\Pr(E|H) = \Pr(E|\neg H)$ .

Indeed, there are *many* logically equivalent ways of writing down  $\mathcal{R}$ , which *syntactically* suggest the salience of different comparisons for the purposes of defining relevance measures of confirmation. While  $\mathcal{R}_1$  may suggest the salience of comparing  $\Pr(H | E)$  and  $\Pr(H)$ ,  $\mathcal{R}_2$  suggests the salience of comparing  $\Pr(E | H)$  and  $\Pr(E | \neg H)$ . The truth is that the relevance criterion is neutral on this question. That is why we shouldn't put too much stock on the syntactical form of measures. Rather, what matters is ordinal or comparative structure imposed by the measures. Another way of making this same kind of point is to note that the likelihood-ratio (and  $KO$ ) can be expressed solely as a function of  $\Pr(H | E)$  and  $\Pr(H)$ :

$$\frac{\Pr(E | H)}{\Pr(E | \neg H)} = \frac{(1 - \Pr(H)) \cdot \Pr(H | E)}{(1 - \Pr(H | E)) \cdot \Pr(H)}$$

In this sense, the  $KO$  measures *are* measures of “the degree to which  $E$  raises the probability of  $H$ ”. In fact, *all* relevance measures are measures of *that*. So, there is a false dichotomy here. The real dichotomy is between relevance measures of *inferential strength* (i.e., those that generalize entailment in the sense captured by the deductive evidence desideratum), and relevance measures that are *not* measures of inferential strength.

It is clear that Fitelson is after an explication of (relevant) *inferential strength* and *not* some *other* kind of relevance confirmation. So, I see no need for the author to digress into their discussion of what they are calling “*size of probability increase*”. This discussion is not relevant to Fitelson's work, and so it is a tangent that should be omitted. I wish I could see a way to make this relevant to the ultimate point of the paper. It certainly is important to note that Fitelson's desiderata imply that the prior probability of the hypothesis should be *irrelevant* to the strength of the argument in cases where there is entailment or refutation. But, this seems obvious (since it is obvious that the *logical* goodness of a valid argument should not depend on how confident one is in the premises or the conclusion of said argument). So, it seems clear that it is no objection to Fitelson's measure(s) that it fails to be sensitive to the priors of

---

<sup>1</sup>Strictly speaking, these are not logically equivalent, since judgments of independence understood as  $\Pr(H|E) = \Pr(H)$  vs  $\Pr(E|H) = \Pr(E|\neg H)$  can diverge in cases where  $E$  and/or  $H$  are not logically contingent (at least, on a Kolmogorovian theory of probability). But, the author restricts their discussion to these cases anyway, so there is no loss of generality here. I will return to the “logical contingency of  $H, E$ ” issue, below, as it arises elsewhere in the author's discussion.

conclusions of valid arguments. Indeed, that is a *virtue* of a measure of *inferential strength*, which is what Fitelson is after (as the author notes). So, I think this entire discussion of “*size of probability increase*” should be omitted from the paper. I’ll say more about this, below.

4. The two “negligible worries” on page 3 are really rather trivial and should also (ultimately) be omitted. First, Fitelson is well aware that the choice of positive/negative signs for  $c$  in cases of confirmation/disconfirmation is just a useful convention and has no probative value. In Fitelson’s dissertation, he notes that the technically picky way to state the relevance constraint on measures  $c$  is to say that there should exist some order-preserving function  $f$  such that  $f(c)$  is positive/negative in cases of confirmation/disconfirmation. So much for the first “negligible worry”. Second, the cases in which  $E$  and/or  $H$  are not logically contingent are cases in which *many* relevance measures – not just the *KO* measures – will be *undefined* (assuming Kolmogorov’s theory of probability). As such, it is no special problem for Fitelson’s desiderata that they ignore such cases. Moreover, confirmation theory is almost never applied to cases in which  $E$  and/or  $H$  are non-contingent. Such cases are just uninteresting from an inductive-logical point of view. Moreover, even deductive logic has problems with the paradoxes of entailment. So, restricting our discussion to non-paradox-of-entailment cases seems reasonable. Thus, there is no real worry here (at least, no worry that is special to Fitelson and his work) and so the second “negligible worry” should also be omitted.
5. Section 2 is all about the “*size of probability increase*” vs *inferential strength* distinction. As I explain above, I don’t see the ultimate relevance of this to Fitelson’s work. As the author himself concedes:

There is, of course, nothing to be said against the . . . project of searching for a quantitative generalisation of deductive entailment, i.e., a measure that gives the strength of the argument from evidence to hypothesis. This is a respectable and well-defined explicational task, and pro- visos (a) and (e) seem to be unavoidable within this project.

I couldn’t have said this better myself! The truth is that Fitelson never claims to be giving “the one true measure of confirmation”. Rather, Fitelson aims to explicate (to use the author’s language) “degree of inferential strength”. Note that most of the author’s quotes come from Fitelson’s entry on *Inductive Logic*, which is all about explicating logical argument strength. And, to this end, Fitelson has a stronger case for the *KO* measures (taking into account the totality of his work) than the author seems to realize (on which more below). It is simply not a criticism of Fitelson to point out that his explicatum is inadequate for an explicandum that he is not trying to explicate! So, I think this section should be omitted. Let’s criticize Fitelson on his own terms. Let’s not tell him what he “should be trying to explicate” (whatever that means). Let’s try to show that there is

a residual explicatory burden for Fitelson – given the explicandum he has in mind, which is clearly *not* “*size of probability increase*”. I take it that the main point of the author’s discussion is to try to do just this. As such, I say omit section 2 (and the earlier stuff which suggests similar “objections”) and just focus on the *internal* objection to Fitelson’s explicatory project. That objection is articulated well in section 3. I think the entire paper should be based on section 3 and the parts of the introduction that don’t slide into the explicational irrelevancies of section 2.

6. OK, that finally brings us to section 3, which is *the* salient part of this paper. In this section, the author correctly points out that there are other measures (which are *not* ordinally equivalent to the *KO* measures) that satisfy Fitelson’s deductive evidence desideratum. This calls into question the strength of the claim to uniqueness that Fitelson is entitled to. As I mentioned above, Fitelson never claims *mathematical* uniqueness of the *KO*-measures. Rather, he claims *historical* uniqueness (relative to the class of confirmation measures that had been proposed up until 2004). Fitelson was aware that there were other relevance measures satisfying his deductive evidence desideratum, but he was not aware of the *particular* measure that the author cleverly designs in this section.

The author motivates their interesting measure as a measure of “exhaustion of probabilistic leeway”. This is an interesting idea, which merits further scrutiny. I would suggest that the author expand a bit on this concept (as *explicandum*). It is important to note that this concept “exhaustion of probabilistic leeway” is not the same as “*size of probability increase*”, which is another reason to omit the discussion of (measures of) the latter in favor of the former (which the author ends-up pitting against the *KO*-measures anyway). I guess I don’t have too much to say about this new explicandum the author has in mind. I suppose it has some psychological validity, but, whatever it is, I do not think it is consonant with the *inferential strength* concept that Fitelson aims to explicate. And, as a result, I do not think the author’s (legitimate, internal) criticism of Fitelson in §3 ultimately succeeds. Before explaining why, let me discuss a simpler measure that satisfies Fitelson’s deductive evidence desideratum:

$$\lambda(H, E) = \frac{\Pr(H | E)}{\Pr(H | \neg E)}$$

It is easy to see that  $\lambda$  is a relevance measure that satisfies Fitelson’s deductive evidence desideratum. But,  $\lambda$  violates other desiderata (now widely accepted) for measures of inferential strength that Fitelson has discussed elsewhere. In particular,  $\lambda$  violates the following desideratum:

(†) If  $\Pr(H | E_1) > \Pr(H | E_2)$ , then  $\mathfrak{c}(H, E_1) > \mathfrak{c}(H, E_2)$ .

And,  $\lambda$  violates the evidence symmetry desideratum of Eells & Fitelson:

(ES) There exists an  $E$  and an  $H$  such that  $\mathfrak{c}(H, E) \neq -\mathfrak{c}(H, \neg E)$ .

Happily, the author's measure is much more clever than  $\lambda$ . The author defines a piecewise (and therefore discontinuous<sup>2</sup>) confirmation measure:

$$y(H, E) = \begin{cases} \frac{\Pr(H | E) - \Pr(H)}{1 - \Pr(H)} & \text{if } \Pr(H | E) \geq \Pr(H). \\ \frac{\Pr(H | E) - \Pr(H)}{\Pr(H)} & \text{if } \Pr(H | E) < \Pr(H). \end{cases}$$

This  $y$  is a very clever and interesting measure. And, to my knowledge,  $y$  is new on the scene. As the author correctly shows,  $y$  satisfies Fitelson's deductive evidence desideratum, and  $y$  is not ordinally equivalent to the  $KO$ -measures. This much does not make it any more interesting than  $\lambda$ , which was known already to Fitelson. But, the author points out many other important properties of  $y$  which make it much more interesting and serious as a candidate measure of degree of inferential strength and therefore a serious challenge to the  $KO$ -measures as explicata of the kind Fitelson seeks. Unfortunately, I think the author's discussion of  $y$  comes up a little bit short, and so, in the end, does  $y$  itself, as a measure of inferential strength (I cannot comment on the adequacy of the measure  $y$  as an explicatum for "exhaustion of probabilistic leeway" because (a) I am not sure what this explicandum is supposed to be, and (b) besides it wouldn't be relevant to a paper that aims to criticize *Fitelson's* project).

The author does not discuss desideratum ( $\dagger$ ), but it is easy to show that  $y$  also satisfies ( $\dagger$ ) as well (the author should discuss this in their paper). To his great credit, the author does discuss (ES), and he rightly explains that  $y$  satisfies (ES). In these ways,  $y$  is certainly superior to  $\lambda$  or any other non- $KO$ -relevance measure of inferential strength of which I am aware.

The author also rightly and commendably explains that  $y$  satisfies the hypothesis symmetry desideratum (HS) of Eells & Fitelson, which is:

$$(HS) \text{ For all } E \text{ and } H, c(H, E) = -c(\neg H, E).$$

Moreover, the author is to be further commended for discussing two other important desiderata for inferential strength that have been discussed by Fitelson and others. It is in their discussion of these two additional desiderata that I think the author's critique founders.

The first additional desideratum the author discusses has to do with the "irrelevant conjunction" problem. I won't get into all the details here. Rather, I'll just state the desideratum the author discusses here:

---

<sup>2</sup>Fitelson, Milne, Good, Heckerman, and others in this literature tend to restrict their attention to continuous, non-piecewise measures of confirmation. Indeed, Milne, Good, and Heckerman explicitly assume continuity as a premise in their "uniqueness proofs". Fitelson has always been suspicious of such assumptions (which is why he has never tried to give a mathematical uniqueness proof like these other authors have), and so he welcomes this clever maneuver of the author. Note: all the historical measures to date have been continuous. [My conjecture is that this has been due mainly to historical prejudice, mathematical laziness, and lack of imagination.]

(‡) For all  $H, E$ , and  $X$ , if  $H \models E$ , then  $c(H \& X, E) < c(H, E)$ .

The author shows that  $\gamma$  satisfies (‡). While it is true (and important) that  $\gamma$  satisfies (‡),  $\gamma$  violates a more general and important “irrelevant conjunction” desideratum discussed by Fitelson (and Hawthorne & Fitelson). Again, I won’t get into the details of this, but the desideratum is:

(I) For all  $H, E$ , and  $X$ , if  $X$  is confirmationally irrelevant (in the standard Bayesian sense) to  $H, E$ , and  $H \& E$ , then  $c(H \& X, E) < c(H, E)$ .

As Fitelson (and Hawthorne & Fitelson) explain, it is (I) and not (‡) that really captures the behavior necessary for a measure  $c$  to handle the Bayesian “irrelevant conjunction” problem. It is easy to see that  $\gamma$  violates (I). If  $E$  disconfirms  $H$  and  $X$  is an “irrelevant conjunct” in the strong sense defined in (I), then  $E$  also disconfirms  $H \& X$  (see Fitelson’s irrelevant conjunction paper for a proof of this). So, in such cases:

$$\begin{aligned} \gamma(H \& X, E) &= \frac{\Pr(H \& X | E) - \Pr(H \& X)}{\Pr(H \& X)} \\ &= r(H \& X, E) - 1 \\ &= r(H, E) - 1 \\ &= \gamma(H, E) \end{aligned}$$

where  $r(H, E) = \frac{\Pr(H|E)}{\Pr(H)}$  is the ratio measure of confirmation, which, as Fitelson (and Hawthorne & Fitelson) explain, violates (I) in *all* cases, since *whenever*  $X$  is an irrelevant conjunct,  $r(H \& X, E) = r(H, E)$ . The author’s  $\gamma$  is only better than  $r$  in this regard when  $E$  confirms  $H$ . When  $E$  disconfirms  $H$ , however, the author’s  $\gamma$  is ordinally equivalent to  $r$ , and so inherits all of its faults in these cases. There is no reason why principle (I) should only apply to confirmatory cases (surely, that would be an ungrounded asymmetry). Unfortunately, this same problem plagues the author’s discussion of the second additional desideratum.

The second additional desideratum is the following (of Eells & Fitelson):

(CS) There exists an  $E$  and an  $H$  such that  $c(H, E) \neq c(E, H)$ .

As the author shows,  $\gamma$  satisfies (CS). In this sense,  $\gamma$  is better than  $r$ , which violates (CS). However, as in the case of (I), above, this advantage over  $r$  is seen only in cases where  $E$  confirms  $H$ . If  $E$  disconfirms  $H$ , then  $\gamma$  is ordinally equivalent to  $r$ , and so inherits its shortcomings, including the violation of (CS). In other words,  $\gamma$  violates the equally compelling:

(CS’) There exists an  $E$  and an  $H$  such that  $c(H, E) \neq c(E, H)$  **and**  $c(H, E) < 0$ .

In fact, the original (unpublished) example that motivated (CS) was invented by Ellery Eells, and it involved a case of *disconfirmation*, not confirmation. Again there is no reason why (CS) should only make sense for cases of confirmation (where would *that* asymmetry come from?).

What this shows is that the author's piecewise definition of  $\gamma$ , while clever and useful for avoiding various problems seen in non-*KO*-candidate relevance measures of inferential strength, also brings along with it some unwanted asymmetries between cases of confirmation vs disconfirmation. In cases of confirmation, the measure behaves like a deductively-corrected difference measure (which is a *good* thing). But, in cases of disconfirmation, the measure behaves just like the ratio measure (which is a *bad* thing). So, in the end, I conclude that the author's proposed measure is not a viable measure of inferential strength, and so poses no serious threat to the *KO*-measures as the best (currently known) candidate relevance measures of inferential strength.

7. As I mentioned above, the author does not say anything about how their measure handles *background knowledge*. Typically, measures of confirmation are defined with three arguments:  $E$ ,  $H$ , and  $K$  (a background corpus). I assume (following standard practice) the author would define:

$$\gamma(H, E | K) = \begin{cases} \frac{\Pr(H | E \& K) - \Pr(H | K)}{1 - \Pr(H | K)} & \text{if } \Pr(H | E \& K) \geq \Pr(H | K). \\ \frac{\Pr(H | E \& K) - \Pr(H | K)}{\Pr(H | K)} & \text{if } \Pr(H | E \& K) < \Pr(H | K). \end{cases}$$

Then, we must address the adequacy of  $\gamma(H, E | K)$  *in general*, and not just in cases where  $K = \top$ . Here, I think new challenges for  $\gamma$  arise. For instance, Fitelson discusses various desiderata concerning *independent evidence*. Fitelson defines the notion of confirmational independence, as:

*$E_1$  is confirmationally independent of  $E_2$  regarding  $H$  according to  $\mathfrak{c}$  iff  $\mathfrak{c}(H, E_1 | E_2) = \mathfrak{c}(H, E_1 | \top)$ . [If  $\mathfrak{c}(H, E_1 | E_2) = \mathfrak{c}(H, E_1 | \top)$  and  $\mathfrak{c}(H, E_2 | E_1) = \mathfrak{c}(H, E_2 | \top)$ , then  $E_1$  and  $E_2$  are said to be *mutually confirmationally independent* regarding  $H$ .]*

As Fitelson shows, the *KO*-measures (as well as the non-*KO*-measures  $d$  and  $r$ ) have some desirable independence properties. For instance:

- If  $\mathfrak{c}(H, E_1 | E_2) = \mathfrak{c}(H, E_1 | \top)$ , then  $\mathfrak{c}(H, E_2 | E_1) = \mathfrak{c}(H, E_2 | \top)$ . [This says that confirmational independence is *symmetric* in  $E_1$  and  $E_2$ .]
- If  $\mathfrak{c}(H, E_1 | E_2) = \mathfrak{c}(H, E_1 | \top)$ , then there exists an function  $f$  such that  $\mathfrak{c}(H, E_1 \& E_2 | \top) = f(\mathfrak{c}(H, E_1 | \top), \mathfrak{c}(H, E_2 | \top))$ . [The joint support of two confirmationally independent pieces of evidence should be a function of their individual degrees of support.]

As defined,  $\gamma(H, E | K)$  will violate both of these desiderata concerning independent evidence. What's worse, we have the following *triviality result* for  $\gamma$  which stems from its peculiar piecewise (discontinuous) definition:

- If  $E_1$  confirms  $H$  and  $E_2$  disconfirms  $H$ , then  $E_1$  and  $E_2$  *cannot* be mutually confirmationally independent regarding  $H$  according to  $\gamma$ .

This is a very undesirable result caused by the discontinuous way in which  $\gamma$  handles confirmatory vs disconfirmatory evidence. I am sure there are still other desiderata which trade on how a measure handles perturbations to background knowledge that would plague  $\gamma$  because of its discontinuous definition, but I don't have time right now to work through all of the ramifications of this. In any case, considerations involving the handling of background knowledge are crucial for any ultimate adjudication.

To sum up: This is a clearly written, well researched, and well argued paper. But, I think much of the author's discussion is irrelevant to Fitelson's explicatory project. Section 3 is the only really salient section. There, the author constructs a very clever alternative to Fitelson's proposed explicatum for (relevant) inferential strength, and he shows that it satisfies many important desiderata for candidate measures of inferential strength. But, in the end, the measure fails to be an adequate explicatum in this sense, because (a) in cases of disconfirmation it behaves exactly like the ratio measure, which has been shown by Fitelson, Eells & Fitelson, and Hawthorne & Fitelson to be inadequate as an explicatum of inferential strength, and (b) in the more general setting which takes account of how measures handle perturbations of background knowledge (such questions are simply ignored by the author, but they are crucial for a final judgment about measures) still further problems with  $\gamma$  are seen to arise (*e.g.*, in  $\gamma$ 's handling of *independent evidence*). If the author can find a way to persuasively defend their measure (or some modification of it) against these charges, then I would be happy to see a revised version of the paper that does so. If not, then I'm afraid this interesting and clever attempted critique of Fitelson will fail.<sup>3</sup>

---

**Postscript on Carnap.** In the first edition of *LFP*, Carnap uses a relevance ("confirmation as increase in firmness") measure that is *not* equivalent to  $d(H, E) = \Pr(H | E) - \Pr(H)$ . Indeed, the relevance measure  $r(H, E) = \Pr(H \& E) - \Pr(H) \cdot \Pr(E)$  Carnap uses in the first edition of *LFP* doesn't even satisfy ( $\dagger$ ), which is widely accepted as a desideratum both for inferential strength and "*size of probability increase*"! So, I think it is clear that in the first edition of the *LFP*, Carnap did *not* equate "confirmation as increase in firmness" and "*size of probability increase*". In *LFP*<sub>2</sub>, he added a preface in which he *suggests*  $d$  as a *possible* measure of "confirmation as increase in firmness". But, he does *not* offer a *theory* of "confirmation as increase in firmness", nor does he offer a theory of "*size of probability increase*" (whatever *that* is). Kemeny & Oppenheim, I take it, were (among other things) trying to fill this gap in Carnap's work. So, as they stand, I think the author's remarks on Carnap are misleading. I recommend *The Popper-Carnap Controversy* by Michalos (Nijhoff, 1971) for the salient history.

---

<sup>3</sup>I cannot comment on the adequacy of the proposed measure as an explicatum for the new explicandum "exhaustion of probabilistic leeway" which is briefly discussed by the author. I'm not sure what this concept is, and I am even less sure as to what its relevance might be to traditional problems in inductive logic and/or Bayesian confirmation theory. In any event, a positive explicatory project for the concept "exhaustion of probabilistic leeway" would be a much different project than the one undertaken in the current manuscript, which is being sold as a critique of Fitelson's explicatory project. That would be a different paper altogether, I think.