Confirmation Measures and Sensitivity

Olav B. Vassend*†

Stanley Stevens draws a useful distinction among ordinal scales, interval scales, and ratio scales. Most recent discussions of confirmation measures have proceeded on the ordinal level of analysis. In this article, I give a more quantitative analysis. In particular, I show that the requirement that our desired confirmation measure be at least an interval measure naturally yields necessary conditions that jointly entail the log-likelihood measure. Thus, I conclude that the log-likelihood measure is the only good candidate interval (or ratio) measure.

1. Introduction. Suppose our preferred confirmation measure, c, outputs the numbers $c(H_1, E) = 0.1$, $c(H_2, E) = 0.2$, $c(H_3, E) = 0.3$, $c(H_4, E) = 50$ for hypotheses H_1, H_2, H_3, and H_4, given evidence E. It is natural to want to say that H_1 and H_2 are confirmed to roughly the same (low) degree by E and that H_4 is confirmed by E to a much higher degree than either H_1 or H_2. We might also want to say that the difference in confirmation conferred by E on H_1 as opposed to on H_2 is the same as the difference in confirmation conferred by E on H_2 as opposed to on H_3. If we make any of the preceding assertions, we are implicitly relying on the assumption that it is legitimate to interpret the differences between the numbers outputted by measure c. In other words, we are assuming that c is at least an *interval measure* in the terminology of Stevens (1946). In this article I show how the preceding assumption, when properly spelled out, places stringent requirements on c that considerably narrow down the field of potential confirmation measures. In fact, I show that only the log-likelihood measure meets the requirements. My argument does not, however, establish that the log-likelihood measure is the true measure of confirmation; the argument only shows that the log

*To contact the author, please write to: Department of Philosophy, University of Wisconsin–Madison, 600 North Park Street, Madison, WI 53706; e-mail: vassend@wisc.edu.

likelihood is the only candidate interval or ratio measure. This leaves it open that there is no adequate confirmation measure that is at least an interval measure.

I start by laying out my background assumptions in section 2. In section 3, I make the requirements on c more precise. In section 4, I show how these requirements entail that c is the log-likelihood measure. In section 5, I discuss the implications of the argument and consider a couple of objections.

2. Background Assumptions. According to a criterion of confirmation universally agreed on among Bayesians, E confirms H just in case $\Pr(H|E) > \Pr(H)$.[1] Although this criterion suffices to answer the binary question whether E confirms H, it does not answer the quantitative question whether E confirms H to a high degree, nor does it answer the comparative question which of two hypotheses is confirmed more by E.[2] In order to answer either of the preceding types of questions, one needs a confirmation measure that quantifies the degree to which E confirms (or disconfirms) H. What follows is a small sample of the measures that have been offered in the literature:

The plain ratio measure, $r(H, E) = \Pr(H|E)/\Pr(H)$.
The log-ratio measure, $lr(H, E) = \log r(H, E)$.
The difference measure, $d(H, E) = \Pr(H|E) - \Pr(H)$.
The log-likelihood measure, $l(H, E) = \log(\Pr(E|H)/\Pr(E|\neg H))$.
The alternative difference measure, $s(H, E) = \Pr(H|E) - \Pr(H|\neg E)$.[3]

Since Bayesians analyze confirmation in terms of probability, and since the probability distribution over the algebra generated by H and E is determined by $\Pr(H|E)$, $\Pr(H)$, and $\Pr(E)$, it has become standard to assume that any confirmation measure can be expressed as a function of $\Pr(H|E)$, $\Pr(H)$, and $\Pr(E)$. The preceding assumption is essentially the requirement that Crupi, Chater, and Tentori (2013) call "formality." A strong case can, however, be made for not allowing our measure of confirmation to depend on $\Pr(E)$. As Atkinson (2009) points out, if we let $c(H, E)$ be a function of $\Pr(E)$, then $c(H, E)$ can change even if we add to E a piece of irrelevant "evidence" E' that is probabilistically independent of H and E and of their conjunction. To see this, suppose that $c(H, E) = f(\Pr(H), \Pr(H|E), \Pr(E))$. Let E' be any proposition whatsoever that is independent of H, E, and

1. *Disconfirmation* happens when the inequality sign is reversed, and when there is an equality sign we have *confirmation neutrality*.

2. Carnap (1962) was the first philosopher to draw the distinction among these three questions.

3. This measure is also sometimes called the Joyce-Christensen measure, after Joyce (1999) and Christensen (1999).

$H\&E$.[4] Then $c(H, E\&E') = f(\Pr(H), \Pr(H|E\&E'), \Pr(E\&E')) = f(\Pr(H),$ $\Pr(H|E), \Pr(E)\Pr(E'))$. If f depends on the third argument, we can find some probability function Pr such that $f(\Pr(H), \Pr(H|E), \Pr(E)\Pr(E')) \neq$ $f(\Pr(H), \Pr(H|E), \Pr(E))$ and thus such that $c(H, E\&E') \neq c(H, E)$. However, this is clearly counterintuitive, since E' is probabilistically independent of H and E and therefore should not have any impact on the confirmation of H. So we conclude that f should not depend on $\Pr(E)$.

Since I find the preceding argument convincing, I assume that the confirmation measure we are looking for is of the following form: $c(H, E) =$ $f(\Pr(H), \Pr(H|E))$. Since there is no a priori restriction on what credences an agent may have except that these credences must lie somewhere in the interval $[0, 1]$, I assume that f is defined on all of $[0, 1] * [0, 1]$. Note that, as Huber (2008) points out, this is not the same as assuming that any particular probability distribution $\Pr(*)$ is continuous.

The preceding two assumptions are summed up in the following requirement:

> **Strong Formality** (SF). Any confirmation measure is of the following form: $c(H, E) = f(\Pr(H), \Pr(H|E))$, where f is a function defined on all of $[0, 1] * [0, 1]$.

It should be noted that SF excludes some of the confirmation measures that have been offered in the literature.[5] I briefly address lingering objections to SF in section 5. Finally, I also adopt the following convention:

> **Confirmation Convention** (CC).

$$c(H, E) : \begin{cases} > 0 & \text{if } \Pr(H|E) > \Pr(H), \\ = 0 & \text{if } \Pr(H|E) = \Pr(H), \\ < 0 & \text{if } \Pr(H|E) < \Pr(H). \end{cases}$$

CC is sometimes taken to be part of the definition of what a confirmation measure is (e.g., by Fitelson 2001). Although I think it is a mistake to think of CC in this way, I adopt CC in this article for convenience. CC has the role of setting 0 as the number that signifies confirmation neutrality.

3. The Main Requirement on c. Suppose we witness a coin being flipped 10 times, and our task is to assign a credence to the proposition that the coin comes up heads on the eleventh flip. If we do not in advance know anything about the coin's bias, it is reasonable to guess that the coin will come up

---

4. I am of course assuming here that H and E are fixed.

5. In particular, the alternative difference measure.

heads with probability $H/10$ on the eleventh flip, where H is the number of times the coin comes up heads in the 10 initial flips.[6] In making this guess, we are setting our credence in the coin landing heads equal to the observed frequency of heads. This move is reasonable since the law of large numbers guarantees that the observed frequency of heads converges in probability to the coin's actual bias. The observed frequency of heads does not necessarily equal the coin's bias after just 10 flips, however. In fact, statistics tells us that the confidence interval around the observed frequency can be approximated by $\hat{p} \pm z[(1/n)\hat{p}(1 - \hat{p})]^{1/2}$, where $\hat{p}$ is the observed frequency, n is the sample size (in this case, 10 coin flips), and z is determined by our desired confidence level.

For example, suppose we witness four heads in 10 coin flips and we set our confidence level to 95%. In that case, $z = 1.96$, $\hat{p} = 0.4$, and the calculated confidence interval is approximately [0.1, 0.7]. Clearly, the confidence interval in this case is rather large. Given our evidence, we can do no better than to estimate the coin's bias as 0.4. However, we also need to realize that if the 10 flips were repeated, we would probably end up with a slightly different value for $\hat{p}$: we should acknowledge that credences are bound to vary with our varying evidence.

The above example illustrates one way that variability can sneak into our credences: if our credence is calibrated to frequency data, then our credence inherits the variability intrinsic to the frequency data. However, even if we set our credence by other means than frequency data, we must admit that rational credences are intrinsically somewhat variable. For example, if the sky looks ominous and I guess that there is a 75% chance that it is going to rain (or perhaps my betting behavior reveals that this is my credence that it is going to rain), I must concede that another agent whose credence (or revealed credence) is 74% or 76% is just as rational as I am: I do not have either the evidence or the expertise to discriminate between these credences. And even if I do have good evidence as well as expertise, I must admit that I am almost never in a position where I have all the evidence, and had I been provided with somewhat different evidence, I would have ended up with a somewhat different credence.

That our credences are variable is a fact of life that any rational agent must face squarely. It is not hard to see that this fact also affects Bayesian confirmation theory. Bayesian confirmation measures are defined in terms of credences and are therefore infected by the variability inherent in credences. If Bayesian confirmation measures are necessarily affected by variable credences, I contend that we should want a confirmation measure that is affected by such variability in a systematic and predictable way. We should want this even if we only care about the ordinal properties of con-

6. This assumes $0 < H < 10$.

firmation measures. Suppose, for instance, that our confirmation measure is very sensitive to minor variations in the prior or the posterior. In that case, if we find out that $c(H, E) > c(H', E')$, we cannot necessarily be confident that H truly is better confirmed by E than H' is by E' because a small variation in our credence in H or H' might well flip the inequality sign so that we instead have $c(H, E) < c(H', E')$. In order to be confident that $c(H, E)$ really is better confirmed than $c(H', E')$, we need to be assured that the inequality sign is stable. Now, we can be assured that the inequality is stable as long as $c(H, E) - c(H', E')$ is of "significant size." But in order for us to be able to determine that $c(H, E) - c(H', E')$ is "of significant size," we need to be able to draw meaningful and robust conclusions from this difference.

Thus, even if we are primarily interested in the ordinal ranking of evidence-hypothesis pairs provided by c, we still want to be able to draw conclusions from the difference $c(H, E) - c(H', E')$. However, if c is very sensitive to small variations in the priors or posteriors of H and H', then the quantity $c(H, E) - c(H', E')$ is unstable: it could easily have been different, since our priors or posteriors could easily have been slightly different (e.g., if we calibrated our priors to frequency data). We are therefore only justified in interpreting the difference $c(H, E) - c(H', E')$ if c is relatively insensitive to small variations in the priors and posteriors.

Suppose, moreover, that slight variations in small priors (or posteriors) have a larger effect on c's output than do slight variations in larger priors. Then we cannot compare the quantity $c(H, E) - c(H', E)$ to the quantity $c(H'', E) - c(H', E)$ unless our prior credences in H'' and H are approximately the same. In order for us to be able to compare $c(H, E) - c(H', E)$ to $c(H'', E) - c(H', E)$ in cases in which our prior credences in H'' and H are very different, we need c to be uniformly insensitive to small variations in the prior (and the posterior). We can sum up the preceding two remarks as follows:

> **Main Requirement** (MR). We are justified in interpreting and drawing conclusions from the quantity $c(H, E) - c(H', E')$ only if c is uniformly insensitive to small variations in $\Pr(H)$ and $\Pr(H|E)$.

As it stands, MR is vague. What counts as a small variation in a credence? Moreover, what does it mean, concretely, for c to be uniformly insensitive to such variations? To get a better handle on these questions, let us formalize the important quantities that occur in MR. Following SF, we are assuming that c is of the form $c(H, E) = f(\Pr(H), \Pr(H|E))$. For simplicity, let us put $\Pr(H) = x$ and $\Pr(H|E) = y$, so that $c = f(x, y)$. According to MR, we require that f be uniformly insensitive to small variations in x and y. I will use $v(p, \varepsilon)$ to capture the notion of a small variation in the probability p, where ε is a parameter denoting the size of the variation. Moreover, I

will use $\Delta_\varepsilon^x c(x, y)$ to denote the variation in c that results from a variation of size ε about x. That is to say,

$$\Delta_\varepsilon^x c(x, y) = f(v(x, \varepsilon), y) - f(x, y). \tag{1}$$

Similarly, I will use $\Delta_\varepsilon^y c(x, y)$ to denote the variation in c that results from a variation of size ε about y. Thus,

$$\Delta_\varepsilon^y c(x, y) = f(x, v(y, \varepsilon)) - f(x, y). \tag{2}$$

The next step is to get a better grip on MR by investigating the terms that occur in (1) and (2). In sections 3.1–3.3, that is what I do.

3.1. What Is Uniform Insensitivity?   First, the demand that c be uniformly insensitive to variations in the prior and the posterior now has an easy formal counterpart: it is simply the demand that for different values x_1, x_2, y_1, and y_2 of x and y, we have $\Delta_\varepsilon^{x_1} c(x_1, y_1) = \Delta_\varepsilon^{x_2} c(x_2, y_2) = \Delta_\varepsilon^{x_2} c(x_2, y_1)$, and so on, and $\Delta_\varepsilon^{y_1} c(x_1, y_1) = \Delta_\varepsilon^{y_2} c(x_2, y_2) = \Delta_\varepsilon^{y_2} c(x_1, y_2)$, and so on. Thus, across different values of x and y, a small variation in c will mean the same thing. More important, this means that we can consider $\Delta_\varepsilon^x c(x, y)$ as purely a function of ε, and likewise for $\Delta_\varepsilon^y c(x, y)$. From now on, I therefore write

$$g(\varepsilon) := \Delta_\varepsilon^x c(x, y). \tag{3}$$

$$h(\varepsilon) := \Delta_\varepsilon^y c(x, y). \tag{4}$$

In order to figure out what the requirement that c be insensitive to small variations amounts to, we need to figure out how to quantify variations in credences. It is to this question that I now turn.

3.2. What Is a Small Variation in a Credence?   Given a credence x, what counts as a small variation in x? This question turns out to have a more subtle answer than one might expect. Using the notation from equations (1) and (2), what we are looking for is the form of the function $v(x, \varepsilon)$. Perhaps the most natural functional form to consider is the following one: $v(x, \varepsilon) = x + \varepsilon$. On this model, a small variation in the probability x is modeled as the addition of a (small positive or negative) number to x. However, if we consider specific examples, we see that this model is too crude. For example, supposing that $x = 0.5$, we might consider 0.05 a small variation relative to x. But if we consider $x = 0.00001$ instead, then 0.05 is no longer small relative to x; instead, it is now several orders of magnitude bigger.

The above example shows that the additive model cannot be right. An easy fix is to scale the size of the variation with the size of x. In other words, we might suggest the following form for v: $v(x, \varepsilon) = x + x\varepsilon$. This adjustment solves the problem mentioned in the previous paragraph. Ac-

cording to the new v, a variation of size 0.025 about 0.5 is "equal" to a variation of 0.0000005 about 0.00001. In contrast to the previous additive model, $v(x, \varepsilon) = x + x\varepsilon$ is a "multiplicative" model of variability, as we can see by instead writing it in the following form: $v(x, \varepsilon) = x(1 + \varepsilon)$.

However, the multiplicative model, although much better than the additive model, is still insufficient. One problem is purely mathematical. Since $v(x, \varepsilon)$ is supposed to correspond to a small shift in probability, we should require that $0 \leq v(x, \varepsilon) \leq 1$, for all values of x and ε. However, $x + x\varepsilon$ can easily be larger than 1, for example, if $x = 0.9$ and $\varepsilon = 0.2$.[7] The other problem is that $v(x, \varepsilon)$ treats values of x close to 0 very differently from values of x close to 1. For instance, a variation where $\varepsilon = 0.1$ will be scaled to 0.001 when $x = 0.01$. But when $x = 0.99$, the same ε will be scaled to just 0.099. This is very problematic, since for every hypothesis H in which we have a credence of 0.99, there corresponds a hypothesis in which we have a credence of 0.01, namely, $\neg H$. But a small variation in our credence in H is necessarily also a small variation in our credence in $\neg H$, simply because $\Pr(\neg H) = 1 - \Pr(H)$: H and $\neg H$ should therefore be treated symmetrically by v. There is an easy fix to both of the preceding problems: if we scale ε by $x(1 - x)$ instead, then first of all we have $0 \leq x + \varepsilon x(1 - x) \leq 1$, and thus $0 \leq v(x, \varepsilon) \leq 1$. Second of all, H and $\neg H$ are now treated symmetrically. From the preceding considerations, we therefore end up with the following functional form for v: $v(x, \varepsilon) = x + x(1 - x)\varepsilon$.

There is a completely different argument by which we can arrive at the same functional form for v. As I mentioned in the example at the beginning of section 3, credences are sometimes calibrated to frequency data. This is, for example, usually the case if H is a medical hypothesis. Suppose H represents the hypothesis that a person P has disease X, for instance. The rational prior credence in H (before a medical examination has taken place) is then the frequency of observed cases of X in the population from which P is drawn. The frequency of observed cases of X can be modeled as the outcome of a binomial process having mean $\Pr(H)$ and variance proportional to $\Pr(H)(1 - \Pr(H))$. Suppose we observe the frequency $fr(\hat{H})$. Then the estimated variance is proportional to $fr(\hat{H})(1 - fr(\hat{H}))$. The variance is maximal at $fr(\hat{H}) = 0.5$ and decreases as $fr(\hat{H})$ moves closer to 0 or to 1. Arguably, it makes a lot of sense in this case for the variability in one's credence to vary with the variance in the frequency data. But that is exactly what $v(x, \varepsilon) = x + x(1 - x)\varepsilon$ does: it scales credence variability by the variance in the data.

From all the preceding considerations, I conclude that what follows is the most plausible functional form for v:[8]

7. This is also a problem for the additive model.

8. I do not claim to have excluded all other possible functional forms for v.

$$v(x, \varepsilon) = x + x(1 - x)\varepsilon. \tag{5}$$

3.3. Uniform Insensitivity to Small Variations in the Prior and Posterior.   The next step is to understand what insensitivity amounts to. To say that c is insensitive to small variations in the prior or posterior is to say that such variations have a small effect on confirmation: the most natural way to formalize this requirement is in terms of continuity. Since $g(\varepsilon)$ represents the change in confirmation resulting from a change (by ε) in probability, a natural continuity requirement for c would be that g and h should be continuous at 0.

However, continuity is too weak a requirement. Even if a function is continuous, it is still possible for it to be very sensitive to small variations. For instance, the function $f(x) = 1,000,000^x$ is continuous (everywhere) but is at the same time very sensitive to small perturbations of x. Sensitivity to perturbations is most naturally measured by looking at how the derivative behaves. Minimally, we should therefore require that g and h be differentiable at 0. The next natural requirement would be to demand that the derivative of both g and h be bounded by some "small" number. Of course, pursuing such a requirement would require a discussion of what is to count as a "small" number in this context. Since I do not actually need a requirement of this sort in my argument in the next section, I will not pursue a discussion of these issues here. The only upshot from this section is therefore that g and h should be differentiable at 0.

4. The Main Result. Let me summarize where we are. Our desire to be able to draw conclusions from differences in confirmation, that is, from expressions of the form $c(H, E) - c(H', E')$, led us to the requirement that c be uniformly insensitive to small variations in $\Pr(H)$ and $\Pr(H|E)$. In sections 3.1–3.3, I made the various components of this requirement more precise. Putting all these components together, we have

> **Formal Version of the Main Requirement** (MR). We are justified in drawing conclusions from the difference $c(H, E) - c(H', E')$ only if the following conditions are all met:
>
> 1. $f(v(x, \varepsilon), y) - f(x, y) = g(\varepsilon)$, where
> 2. $g(\varepsilon)$ does not depend on either x or y
> 3. $g(\varepsilon)$ is differentiable at 0
> 4. $v(x, \varepsilon) = x + x(1 - x)\varepsilon$
> 5. $f(x, v(y, \varepsilon) - f(x, y) = h(\varepsilon)$, where
> 6. $h(\varepsilon)$ does not depend on either x or y
> 7. $h(\varepsilon)$ is differentiable at 0.

Note that 5–7 are just 1–3 except that they hold for h instead of for g. Note also that MR is essentially epistemic. It says that "we" (i.e., agents interested in confirmation) are only justified in drawing conclusions (of any kind) from $c(H, E) - c(H', E')$ if certain formal conditions are met. These conditions ensure that $c(H, E)$ behaves reasonably well. Together with SF and CC, the conditions in MR entail the log-likelihood measure, as I show next.

Main Result. If MR is true, SF is assumed, and CC is adopted as a convention, then

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

where the identity is unique up to multiplication by a positive number.

Proof. Starting with 1 from MR, we have

$$f(v(x, \varepsilon), y) - f(x, y) = g(\varepsilon). \tag{6}$$

If we divide each side by $x(1 - x)\,\varepsilon$, we get

$$\frac{f(v(x, \varepsilon), y) - f(x, y)}{x(1 - x)\varepsilon} = \frac{g(\varepsilon)}{x(1 - x)\varepsilon}. \tag{7}$$

Next, we let $\varepsilon \to 0$:

$$\lim_{\varepsilon \to 0} \frac{f(v(x, \varepsilon), y) - f(x, y)}{x(1 - x)\varepsilon} = \lim_{\varepsilon \to 0} \frac{g(\varepsilon)}{x(1 - x)\varepsilon}. \tag{8}$$

Since g is differentiable at 0 (from part 3 of MR), the right-hand side of the above equation is just $[1/x(1 - x)]g'(0)$. Since the limit exists on the right-hand side of the equation, it must exit on the left side as well. But the left side is just $(\partial/\partial x)f(x, y)$. We therefore have

$$\frac{\partial}{\partial x}f(x, y) = \frac{1}{x(1 - x)}g'(0). \tag{9}$$

Next, we take the antiderivative of each side of (9) with respect to x. Since g and hence $g'(0)$ does not depend on x (from part 2 of MR), we have

$$f(x, y) = g'(0)(\log x - \log(1 - x)) + C. \tag{10}$$

Here, C is a number that depends on y but not on x. If we perform the above calculations again starting instead with $f(0.5, v(y, \varepsilon)) - f(0.5, y) = h(\varepsilon)$ and using (10), we find that

$$C = h'(0)(\log y - \log(1 - y)) + K. \tag{11}$$

Here, K is just a constant (i.e., it depends on neither x nor y). We therefore have

$$f(x, y) = g'(0)(\log x - \log(1 - x)) + h'(0)(\log y - \log(1 - y)) + K. \tag{12}$$

Now set $x = y = 0.5$. The second part of CC then entails that $K = 0$. Next, set $x = y$. Then CC entails

$$g'(0)(\log x - \log(1 - x)) + h'(0)(\log x - \log(1 - x)) = 0. \tag{13}$$

This in turn entails that $g'(0) = -h'(0)$. Thus, we have

$$f(x, y) = -h'(0)(\log x - \log(1 - x)) + h'(0)(\log y - \log(1 - y)) \tag{14}$$

$$= h'(0)\log \frac{y}{1 - y} \times \frac{1 - x}{x}. \tag{15}$$

Remembering that $x = \Pr(H)$ and $y = \Pr(H|E)$, (14) and (15) together with SF entail

$$c(H, E) = f(\Pr(H), \Pr(H|E)) \tag{16}$$

$$= h'(0)\log \frac{\Pr(H|E)}{1 - \Pr(H|E)} \times \frac{1 - \Pr(H)}{\Pr(H)} \tag{17}$$

$$= h'(0)\log \frac{\Pr(H|E)}{\Pr(H)} \times \frac{\Pr(\neg H)}{\Pr(\neg H|E)} \tag{18}$$

$$= h'(0)\log \frac{\Pr(H|E) \times \Pr(E)}{\Pr(H)} \times \frac{\Pr(\neg H)}{\Pr(\neg H|E) \times \Pr(E)} \tag{19}$$

$$= h'(0)\log \frac{\Pr(E|H)}{\Pr(E|\neg H)}. \tag{20}$$

Finally, CC entails that $h'(0)$ must be a positive number. Thus, $c(H, E) = l$, up to multiplication by a positive number. QED

5. Discussion and Objections.
In the previous section, I showed that MR, SF, and CC jointly entail the log-likelihood confirmation measure, l. The proof entails l up to multiplication by a positive number. That is to say, if $\log(\Pr(E|H)/\Pr(E|\neg H))$ is a legitimate confirmation measure, then so is $a \times \log(\Pr(E|H)/\Pr(E|\neg H))$, for $a > 0$; the argument does not establish that any particular logarithmic base is better than another. In Stevens's (1946) ter-

minology, our measure is apparently a *ratio* measure, meaning that we are justified in interpreting both intervals and ratios between outputs of the measure. Analogously, mass is also a ratio measure since it makes sense to say both that the difference between 2 and 4 kilograms is the same as the difference between 4 and 6 kilograms and that 4 kilograms is twice as big as 2 kilograms. It therefore appears that my conclusion is stronger than what I set out to establish: in the introduction, I said that the goal was to find a confirmation measure that can be interpreted as at least an interval measure. But the proof in the previous section actually establishes that l is a ratio measure under the conditions specified.[9]

The second thing to notice about my argument is that it does not actually establish that the log-likelihood measure is the true confirmation measure. This is because MR merely gives necessary conditions and no sufficient ones. Thus, what my argument shows is really a conditional statement: if there is any interval confirmation measure, then that measure is l. The preceding conditional is, of course, equivalent to the following disjunction: either there is no interval confirmation measure or the only interval confirmation measure is l.

The third and final observation I will make about the argument is that it clearly depends very much on the choice of v. In section 3.2 I considered and rejected two other measures of variability: the additive measure, $v(x, \varepsilon) = x + \varepsilon$, and the multiplicative measure, $v(x, \varepsilon) = x + x\varepsilon$. It is natural to ask what confirmation measures we end up with if we instead use these alternative measures of credence variability. The answer, although I will not show this here, is that the additive measure yields the difference confirmation measure, d, whereas the multiplicative measure yields the log-ratio confirmation measure, lr. We can therefore see that d and lr "embody" defective measures of credence variability: arguably, that is a strike against these measures.

Next, I consider a couple of objections to the argument. First, my argument is obviously only sound if the assumptions in MR are correct. However, the assumptions in MR might remind the reader of assumptions made in Good (1960, 1984) and Milne (1996). These assumptions have been criticized by Fitelson as being "strong and implausible" (2001, 28–29 n. 43) and for having "no intuitive connection to material desiderata for inductive logic" (2006, 506 n. 12).

Why does my argument escape Fitelson's criticisms? How is my argument different from the arguments made by Good and Milne? The answer is that, whereas Good and Milne are not interested in the interval properties of their confirmation measures, and the various mathematical assumptions they make therefore seem unmotivated, all the properties listed in MR arise

9. I owe thanks to a reviewer who persuaded me of this.

naturally out of our wish to have a confirmation measure that is at least an interval measure.

Finally, one may object to some of the other background assumptions I make in section 1. In particular, SF may be accused of being too strong since it excludes the alternative difference measure right off the bat. My reply to this objection is as follows: the argument in section 4 can be carried out without SF, but the resulting analysis does not yield the alternative difference measure or any other recognizable confirmation measure. Thus, even if one rejects SF, one cannot use the type of argument I have given in this article to argue for the alternative difference measure or other standard measures that depend nontrivially on $\Pr(E)$.[10]

6. Conclusion. I have argued that there is a set of conditions that any confirmation measure must meet in order to justifiably be interpreted as an interval measure. Furthermore, I have shown that these necessary conditions, together with an additional plausible assumption and a widely accepted convention, jointly entail the log-likelihood measure. My argument does not show that l is an interval measure, but it does show that it is the only measure that stands the chance of being one. Nor does the argument in this article show that l is the "true" confirmation measure. However, to the extent that we care about our measure's being an interval measure, we should regard the conclusion in this article as favoring l as our preferred measure.

REFERENCES

Atkinson, David. 2009. "Confirmation and Justification: A Commentary on Shogenji's Measure." *Synthese* 184:49–61.
Carnap, Rudolph. 1962. *Logical Foundations of Probability.* 2nd ed. Chicago: University of Chicago Press.
Christensen, David. 1999. "Measuring Confirmation." *Journal of Philosophy* 96:437–61.
Crupi, Vincenzo, Nick Chater, and Katya Tentori. 2013. "New Axioms for Probability and Likelihood Ratio Measures." *British Journal for the Philosophy of Science* 64:189–204.
Fitelson, Branden. 2001. "Studies in Bayesian Confirmation Theory." PhD diss., University of Wisconsin–Madison.
———. 2006. "Logical Foundations of Evidential Support." *Philosophy of Science* 73:500–512.
Good, I. J. 1960. "Weight of Evidence, Corroboration, Explanatory Power, Information and the Utility of Experiments." *Journal of the Royal Statistical Society* B 22:319–31.
———. 1984. "The Best Explicatum for Weight of Evidence." *Journal of Statistical Computation and Simulation* 19:294–99.
Huber, Franz. 2008. "Milne's Argument for the Log-Ratio Measure." *Philosophy of Science* 75:413–20.

10. Such as Carnap's measure, $c(H, E) = \Pr(H\&E) - \Pr(H)P(E)$.

Joyce, James. 1999. *The Foundations of Causal Decision Theory*. Cambridge: Cambridge University Press.

Milne, Peter. 1996. "Log[P(hleb)/P(hlb)] Is the One True Measure of Confirmation." *Philosophy of Science* 63:21–26.

Stevens, Stanley S. 1946. "On the Theory of Scales of Measurement." *Science* 103:577–80.