Focused Correlation and Confirmation

Gregory Wheeler

Artificial Intelligence Center - CENTRIA
Department of Computer Science, Universidade Nova de Lisboa
2829-516 Caparica, Portugal
grw@fct.unl.pt

May 20, 2008
To appear in *The British Journal for the Philosophy of Science.*

Abstract. This essay presents results about a deviation from independence measure called *focused correlation*. This measure explicates the formal relationship between probabilistic dependence of an evidence set and the incremental confirmation of a hypothesis, resolves a basic question underlying Peter Klein and Ted Warfield's 'truth-conduciveness' problem for Bayesian coherentism, and provides a qualified rebuttal to Erik Olsson's claim that there is no informative link between correlation and confirmation. The generality of the result is compared to recent programs in Bayesian epistemology that attempt to link correlation and confirmation by utilizing a conditional evidential independence condition. Several properties of focused correlation are also highlighted.

Keywords: Bayesian epistemology, coherence measures, confirmation theory, measures of deviation from independence.

1 Introduction

Recent results in Bayesian epistemology have reinvigorated debate over the coherence theory of justification, particularly the version outlined by Laurence BonJour (Bonjour 1985). One issue receiving attention is whether the idea behind coherentist justification is tenable, which is understood in this literature as whether an increase of coherence among a collection of beliefs is matched by an increase in the likelihood of those or related beliefs being true. Here critics pounce.

Peter Klein and Ted Warfield, for instance, have argued on measure-theoretic grounds that it is impossible for coherence to increase the likelihood of truth (1994, 1996). They maintain that coherence is not 'truth conducive' in the way the coherence theory requires because there isn't the right sort of 'truth connection' between coherent beliefs and true beliefs. Erik Olsson also doubts that there is an informative link between a measure of coherence and increase in the likelihood of truth (Olsson 2005, 135). The preconditions for Olsson's result—constructed in terms of a 'testimonial system' in which witness reports are associated with a credibility measure and are independent from one another—are restrictive. But he thinks they are charitable to coherence theorists, arguing that ‘it is less plausible or even impossible that there could be an interesting measure of coherence that is truth conducive’ under circumstances in which these conditions are not met (Olsson 2005, 135).

What Bayesians coherence theorists and critics alike are doing is devising various types of *deviation from independence* (*dfi*) measures for sets of binary variables, or proposing methods for inducing a partial ordering on sets of variables, and then evaluating the prospects for such constructions to explain the relationship between coherence and incremental *confirmation*. The reduction of coherentist justification to *dfi*-measures is not direct, however, because coherentist *justification* is a diachronic notion and all Bayesian coherence measures to date, along with the so-called impossibility results, concern a synchronic notion of coherence.[1] Although it is widely assumed in this literature that positive and negative results about synchronic coherence are relevant to a theory of diachronic coherence, the relationship between these two notions of coherence is still to be worked out.

Nevertheless, the relationship between *dfi*-measures and confirmation is worth exploring quite apart from Bayesian coherentism and its critics. The relationship between *dfi* and confirmation is the basic issue behind Klein and Warfield’s counter-example, and also behind Olsson’s impossibility result. In so far as these two arguments suggest that there is no informative relationship between *dfi*-measures and confirmation, the results of this essay are a rebuttal.

The paper presents a measure called *focused correlation* that resolves a general formal question behind truth-conduciveness objections. Focused correlation is the ratio of the degree of correlation among a pair of variables conditioned on a hypothesis to the degree of correlation of the evidence *simpliciter*. The measure may be used as an indicator function for when combining pieces of evidence for a hypothesis improves or degrades the incremental confirmation that the evidence provides separately for that hypothesis, and the measure may be extended to an n-variable conditional *dfi*-measure without loss of generality. Thus this measure specifies the conditions under which combining probabilistically dependent evidence increases the incremental confirmation of a hypothesis.

Focused correlation is a poor candidate for explicating coherentist justification, however. In addition to worries about the relationship between synchronic coherence and diachronic coherence, there are features of correlation measures in general that raise concerns about their unqualified use in Bayesian epistemology. Thus we begin the paper with a discussion of the relationship between correlation and coherence in section 2 and highlight some limitations of the Wayne-Shogenji measure advocated by Tomoji Shogenji as a coherence measure in (Shogenji 1999). Then in section 3 we present focused correlation, which is based upon a conditionalized form of the Wayne-Shogenji measure that was introduced by Wayne Myrvold (Myrvold 1996).

[1] Bonjour repeatedly stresses that coherentist justification is a dynamic concept, not a static one. See pages 144, 153, and 169 of (Bonjour 1985), for instance.

2 Correlation Measures

Correlation is one type of relationship that can hold among a pair of variables, and the Wayne-Shogenji similarity measure is a species of correlation measure. In this section four points are discussed. First the relationship between the Wayne-Shogenji measure and Pearson's product moment correlation coefficient is addressed. Then the Wayne-Shogenji measure and a conditional form of the measure are extended to n-variable *dfi*-measures. Next, four examples are introduced that demonstrate there is no direct dependence between the direction of the Wayne-Shogenji measure and the direction of incremental confirmation. Finally, a popular strategy for exploring indirect relationships between correlation and confirmation via conditional evidential independence assumptions is considered and criticized. These four points are addressed in the following four subsections.

2.1 Standard Covariance and Correlation Measures

There are a variety of relationships that can hold among variables, but one of the most basic is *covariance*. The covariance of two random variables x and y is the average of x minus its mean, multiplied by the average of y minus its mean,

$$Cov(x, y) = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y}).$$

In other words, $Cov(x, y)$ is a measure for how variables vary together.

We may also measure the strength of the *linear* relationship between x and y as the covariance of x and y divided by the product of standard deviations of the two variables,

$$r(x, y) = \frac{Cov(x, y)}{\sqrt{Var(x)Var(y)}},$$

where $r(x, y)$ is Pearson's *product moment correlation coefficient* for x and y, which takes values between -1 and 1.

Suppose now that x and y are binary variables, where x is 1 if A, 0 if $\neg A$, and y is 1 if B, 0 if $\neg B$. Suppose too that there is a joint probability distribution over these variables such that $\Pr(A \cap B), \Pr(A \cap \neg B), \Pr(\neg A \cap B)$, and $\Pr(\neg A \cap \neg B)$ are defined. If the probability model over the data for x and y satisfies both the *mean assumption* and the *normal error assumption*,[2] and takes values in the open interval $(0, 1)$, then we may define a correlation measure on *events*, $Cor(A, B)$, in terms of the correlation coefficient r:

$$Cor(A, B) = r_{x,y} = \frac{\Pr(A \cap B) - \Pr(A) \times \Pr(B)}{\sqrt{\Pr(A) \times \Pr(\neg A) \times \Pr(B) \times \Pr(\neg B)}}.$$

[2] The *mean assumption*: the conditional mean of y given x is an unknown linear function of x; the *error assumption*: (i) errors are normally distributed with mean 0 and a known variance σ^2, and (ii) errors for each observation are independent.

This textbook result says the following about the probabilistic relationship between events over which Pr is defined: if and only if two events A and B are *independent*, i.e., $\Pr(A \cap B) = \Pr(A) \times \Pr(B)$, then $Cor(A, B) = 0$; if and only if there is a *positive correlation* between A and B, i.e., $\Pr(A|B) > \Pr(A)$, then $Cor(A, B) > 0$, where $Cor(A, A) = 1$ is stipulated to be the positive limit; if and only if there is a *negative correlation* between A and B, i.e., $\Pr(A|B) < \Pr(A)$, then $Cor(A, B) < 0$, where $Cor(A, \neg A) = -1$ is stipulated to be the negative limit.

2.2 The Wayne-Shogenji Measure

Rather than express degree of probabilistic independence among events A and B directly in terms of a linear correlation coefficient, $Cor(A, B)$, we may instead express correlation as a *weight function* on probabilistic independence,

$$\Pr(A|B) = \Pr(A) \times S(A, B), \tag{1}$$

where A and B are independent if and only if $S(A, B) = 1$, A and B are positively correlated if and only if $S(A, B) > 1$, and A and B are negatively correlated if and only if $S(A, B) < 1$. The measure $S(A, B)$ is a non-linear measure of the degree of probabilistic dependence between A and B.

Andrew Wayne (1995) proposed to interpret S as a 'similarity measure' and noted that (1) is equivalent to

$$S(A, B) = \frac{\Pr(A|B)}{\Pr(A)} = \frac{\Pr(B|A)}{\Pr(B)} = \frac{\Pr(A \cap B)}{\Pr(A) \times \Pr(B)}. \tag{2}$$

This measure was also proposed by Tomoji Shogenji (1999) as a measure of coherence. Whereas Shogenji keeps the restriction on values of variables to the open unit interval, Wayne relaxes this restriction by stipulating that if either $\Pr(A)$ or $\Pr(B)$ is zero, then $S(A, B) = 1$. On either account the measure $S(A, B)$ takes 0 as a negative limit but is without an upper limit: instead, the degree of positive similarity among variables is a function of the size and prior distribution of the data set.

Both Wayne Myrvold (1996, 662) and Shogenji (1999) extend (2) to bodies of evidence consisting of three or more variables,

$$S(A_1, \ldots, A_n) = \frac{\Pr(A_1 \cap \ldots \cap A_n)}{\Pr(A_1) \times \ldots \times \Pr(A_n)}, \tag{3}$$

and Myrvold also defines a conditionalized form,

$$S(A_1, \cdots, A_n|B) = \frac{\Pr(A_1 \cap \ldots \cap A_n|B)}{\Pr(A_1|B) \times \ldots \times \Pr(A_n|B)}. \tag{4}$$

Despite the non-linearity of S, the quantity measured in (2) is degree of correlation among pairs of random variables. Table 1 displays the relationships between S, Cor, and direction of correlation (+, −, or independent) for two variables.

Table 1. Correlation measures

(−) Correlation	Independence	(+) Correlation
$\Pr(A\|B) < \Pr(A)$	$\Pr(A\|B) = \Pr(A)$	$\Pr(A\|B) > \Pr(A)$
$Cor(A, B) < 0$	$Cor(A, B) = 0$	$Cor(A, B) > 0$
$S(A, B) < 1$	$S(A, B) = 1$	$S(A, B) > 1$

Strictly speaking, (3) and (4) are *not* correlation measures because correlation is a binary relation: it is well known that sets of 3 or more evidence variables may indicate a high degree of dependence whereas pairs of those variables may fail to be correlated. Yet (3) and (4) are *dfi*-measures. I will sometimes speak informally of correlation measures to include their 3- or more evidence variable counterparts.

2.3 Interpreting Correlation Measures

In order to evaluate a correlation measure it is necessary to know the *significance* of the correlation and the *strength* of the dependency between variables. Significance is a measure of reliability of the correlation of x and y. Tests for significance are typically based upon the assumption that error, i.e., the deviation of data from the 'true' linear function, is normally distributed. For Pearson's coefficient r strength is the square of the correlation coefficient, which measures the proportion of common variation in x and y. Thus, correlation measures are sensitive to outlying data: since regression minimizes the sum of the squares of distances of data from a line plotted through that data, a single outlying data point can significantly alter the slope of the plotted regression line. Furthermore, variables may be non-linearly correlated. Thus, applying the measure r may correctly indicate that there is no linear correlation, but may falsely indicate that there is only a weak degree of dependency among the variables. It is therefore critical to examine the data underpinning a correlation measure to check against error, since bare correlation measures alone tell you little about the actual dependency between variables.

The interpretations of the Wayne-Shogenji measure given by Bayesian epistemologsts fail to address either the significance or the strength of correlated beliefs. Instead, they typically assume that a probability distribution is representative of the evidence, or representative of a rational agent's assessment of likelihoods. But this assumption is suspect given the many qualifications that accompany the interpretation of correlation measures in statistics. In so far as evidence is taken to be either representative of a state of affairs or an assessment of likelihood bound by objective criteria rather than simply a measure of an agent's credence, critically assessing strength and significance is unavoidable. Furthermore, once we take account of the objective basis for probability assessment, then the interpretation of probability becomes a central concern to correlation-based theories in formal epistemology.

For the sake of the argument here, assume there is a means to evaluate significance and strength of correlated beliefs. Then an open problem for Bayesian accounts of confirmation, testimony, and coherence is to specify what impact, if any at all, combining coherent evidence has upon confirmation. This is the problem that Klein and Warfeld's counter-example and Olsson's results address, which has (misleadingly) been referred to as the 'truth-connection' problem.

Concerning the measure S, intuitions are divided. Wayne proposed, but did not endorse, interpreting S to represent the diversity of evidence thesis (Howson and Urbach 1989, 114). This thesis holds that the *less* similar pieces of evidence for some hypothesis are to one another then the stronger the support this combined set of evidence would give to that hypothesis. On this view $S(A, B) = 1$ represents that evidence A and B are maximally diverse, and maximally diverse evidence offers more support for an hypothesis than 'narrow' evidence, i.e., when A and B are either positively or negatively correlated.

Shogenji proposed S as an account of epistemic coherence, which holds that the *more* similar evidence is to each other, the greater the support it offers to an hypothesis. On Shogenji's interpretation $S(A, B) = 1$ represents neutral evidence, negatively correlated evidence represents incoherent evidence, and positively correlated evidence represents coherent evidence.

As these clashing intuitions about S might suggest, correlation and confirmation may vary independently. Contra Klein and Warfield and the diversity of evidence thesis, combining correlated confirming evidence may increase confirmation of a conditioning hypothesis. Contra Shogenji, 'incoherent' evidence may increase confirmation of a conditioning hypothesis. Furthermore, both positively correlated evidence and negatively correlated evidence may each decrease incremental confirmation of a conditioning hypothesis. Each possibility is demonstrated in four examples.

But first, some terminology will be helpful for discussing these examples. Following L. Jonathan Cohen (1977) we say that evidence A and B *converge upon* h if $\Pr(h|A \cap B) > \Pr(h|A)$. Furthermore, we say that the ratio of $\Pr(h|A \cap B)$ to $\Pr(h|A)$ is a measure of *incremental convergence* from A to $\{A, B\}$ on h, and that this incremental convergence is *positive* if and only if $\Pr(h|A \cap B)/\Pr(h|A)$ is greater than 1, *negative* iff less than 1, and *neutral* iff 1. Finally, billiard balls numbered 1 through 8 are classified as 'solids', 9 through 15 as 'stripes', and the cue ball is neither numbered, striped, nor solid.[3] See the Appendix for calculations.

[3] All examples are constructed from the feature categories *parity*, *color*, and *number*. Some examples form mutually exclusive classes by treating values of each of these categories as features, e.g., the classes 'odd' and 'even'. Although redefining *values* of a feature as distinct features is not a sound practice in general, this classification scheme does not affect the point under discussion but is adopted both to simplify the examples and to illustrate the power of focused correlation to correctly classify the relationship between correlation and confirmation even when there are logical relationships that hold among the categories.

Example 1. An urn contains the balls numbered 1, 2, 3, 14, and 15. Consider the hypothesis h_1 and two pieces of evidence, A_1 and A_2:

h_1: The drawn ball is the 2 ball.
A_1: The drawn ball is solid.
A_2: The drawn ball is even.

Note the following values: $\Pr(h_1) = 0.2$, $\Pr(A_1) = 0.6$, $\Pr(A_2) = 0.4$, $\Pr(h_1|A_1) = \frac{1}{3}$, $\Pr(h_1|A_2) = 0.5$, and $S(A_1, A_2) \approx 0.833$. So, the evidence set $\{A_1, A_2\}$ is negatively correlated. However, the incremental convergence from A_1 to $\{A_1, A_2\}$ is positive, so it is *not* the case that the evidence set $\{A_1, A_2\}$ offers less confirmation than each piece of evidence alone. Note that

$$\frac{\Pr(h_1|A_1)}{\Pr(h_1)} \approx 1.667, \frac{\Pr(h_1|A_2)}{\Pr(h_1)} = 2.5,$$

but

$$\frac{\Pr(h_1|A_1 \cap A_2)}{\Pr(h_1|A_1)} = 3,$$

yet

$$\frac{\Pr(h_1|A_1 \cap A_2)}{\Pr(h_1|A_2)} = 2.$$

Hence, a negatively correlated evidence set $\{A_1, A_2\}$ can increase the confirmation of h_1. □

Example 2. An urn contains the billiard balls numbered 1, 2, 14, 15, and the cue ball. Consider the hypothesis h_2 and two pieces of evidence, A_3 and A_4:

h_2: The drawn ball is striped.
A_3: The drawn ball is odd.
A_4: The drawn ball is either an even solid or an odd striped.

Note the following values: $\Pr(h_2) = 0.4$, $\Pr(A_3) = 0.4$, $\Pr(A_4) = 0.4$, $\Pr(h_2|A_3) = 0.5$, $\Pr(h_2|A_2) = 0.5$, and $S(A_3, A_4) = 1.25$. So, the evidence set $\{A_3, A_4\}$ is positively correlated. And on this example the incremental convergence from both A_3 and A_4 to $\{A_3, A_4\}$ is positive; the evidence set $\{A_3, A_4\}$ confirms h_2 more than A_3 and A_4 individually, i.e.,

$$\frac{\Pr(h_2|A_3)}{\Pr(h_2)} = 1.25 = \frac{\Pr(h_2|A_4)}{\Pr(h_2)},$$

and

$$\frac{\Pr(h_2|A_3 \cap A_4)}{\Pr(h_2|A_3)} = 2 = \frac{\Pr(h_2|A_3 \cap A_4)}{\Pr(h_2|A_4)}.$$

Hence, a positively correlated evidence set $\{A_3, A_4\}$ can increase the confirmation of h_2. □

Example 3. An urn contains the billiard balls numbered 1, 2, 14, 15, and the cue ball. Consider the hypothesis h_3 and two pieces of evidence, A_5 and A_6:

h_3: The drawn ball is solid.
A_5: The drawn ball is odd.
A_6: The drawn ball is even.

Note the following values: $\Pr(h_3) = 0.4$, $\Pr(A_5) = 0.4$, $\Pr(A_6) = 0.4$, $\Pr(h_3|A_5) = 0.5$, $\Pr(h_3|A_6) = 0.5$, and $S(A_5, A_6) = 0$. So, the evidence set $\{A_5, A_6\}$ is negatively correlated. And on this example the incremental convergence from both A_5 and A_6 to $\{A_5, A_6\}$ is negative; the evidence set $\{A_5, A_6\}$ together does not confirm h_3 more than A_5 and A_6 individually, i.e.,

$$\frac{\Pr(h_3|A_5)}{\Pr(h_3)} = 1.25 = \frac{\Pr(h_3|A_6)}{\Pr(h_3)},$$

and

$$\frac{\Pr(h_3|A_5 \cap A_6)}{\Pr(h_3|A_5)} = 0 = \frac{\Pr(h_3|A_5 \cap A_6)}{\Pr(h_3|A_6)}.$$

Hence, a negatively correlated evidence set $\{A_5, A_6\}$ can fail to increase the confirmation of h_3. □

Example 4. An urn contains the billiard balls numbered 2, 4, 13, 15, and the cue ball. Consider the hypothesis h_4 and two pieces of evidence, A_7 and A_8:

h_4: The drawn ball is even.
A_7: The drawn ball is the 2 ball.
A_8: The drawn ball is solid.

Note the following values: $\Pr(h_4) = 0.4$, $\Pr(A_7) = 0.2$, $\Pr(A_8) = 0.4$, $\Pr(h_4|A_7) = 1$, $\Pr(h_4|A_8) = 1$, and $S(A_7, A_8) = 2.5$. So, the evidence set $\{A_7, A_8\}$ is positively correlated, therefore coherent. But on this example the incremental convergence from both A_7 and A_8 to $\{A_7, A_8\}$ is neutral; the evidence set $\{A_7, A_8\}$ together does not confirm h_4 more than A_7 and A_8 individually, i.e.,

$$\frac{\Pr(h_4|A_7)}{\Pr(h_4)} = 2.5 = \frac{\Pr(h_4|A_8)}{\Pr(h_4)},$$

but

$$\frac{\Pr(h_4|A_7 \cap A_8)}{\Pr(h_4|A_7)} = 1 = \frac{\Pr(h_4|A_7 \cap A_8)}{\Pr(h_4|A_8)}.$$

Hence, a positively correlated evidence set $\{A_7, A_8\}$ can fail to increase the confirmation of h_4. □

Therefore, the interpretation of $S(A, B)$ as either a measure of coherence of $\{A, B\}$, or a measure of this set's diversity is mistaken: $S(A, B)$ is a measure of correlation of A and B, and the confirmation boost of a hypothesis conditioned on the evidence set $\{A, B\}$ is indeterminate given only the information that A and B are correlated.

2.4 Correlation and Evidential Independence

Since correlation and confirmation are not directly related to one another it is natural to consider whether there are indirect relationships between the two. One way to do this is to add conditions to a correlation measure in order to find dependencies between (positively) correlated evidence and an increase in confirmation of an hypothesis. A popular condition among formal coherence theorists is a conditional independence assumption called *evidential independence* (Earman 2000, Bovens and Hartmann 2003, Olsson 2005, Shogenji 2007), which may be illustrated with two evidence variables, A and B, and an hypothesis variable, h.

Evidential Independence (EI): A is evidentially independent of B with respect to h if and only if $\Pr(A|B \cap h) = \Pr(A|h)$, and $\Pr(A|B \cap \overline{h}) = \Pr(A|\overline{h})$.

Proposition 1 collects some facts about EI.

Proposition 1. *Iff* $\Pr(A|B \cap h) = \Pr(A|h)$ *and* $\Pr(A|B \cap \overline{h}) = \Pr(A|\overline{h})$*, then*

(*i.*) $\Pr(A \cap B|h) = \Pr(A|h) \times \Pr(B|h)$,
(*ii.*) $\Pr(A \cap B|\overline{h}) = \Pr(A|\overline{h}) \times \Pr(B|\overline{h})$, *and*
(*iii.*) $\Pr(h|A \cap B) = \dfrac{\Pr(A|h)\Pr(B|h)\Pr(h)}{\Pr(A|h)\Pr(B|h)\Pr(h) + \Pr(A|\overline{h})\Pr(B|\overline{h})\Pr(\overline{h})}$.

EI says that pieces of evidence about h do not influence one another's probability except through their effect on h. Coherence theorists have appealed to EI in various guises to make plain Bonjour's claims about *cognitively spontaneous beliefs*, which are non-inferential beliefs that nevertheless bear some relationship to one another. On Bonjour's account, if the contents of several independent, cognitively spontaneous beliefs 'agree' on a proposition p, then that collection of beliefs generates evidence for p (Bonjour 1985, 148). The challenge for probabilistic theories of coherence then is to explain what 'independence' means in this context, and how a set of independent, cognitively spontaneous beliefs can positively influence the likelihood that those beliefs are true.

Bovens and Hartmann (2003) address the challenge by adopting a witness model whereby a witness *report* is distinguished from the *content* of that report. On this account coherence occurs among the contents of reports, and EI is a condition governing sets of report variables about the conjunction of the contents of those reports. Olsson (2005) also proposes a witness model that distinguishes between reports and the contents of reports. According to Olsson, a Bonjour belief system is a set of pairs, $T = \{\langle Rep(p), p\rangle_1, \ldots, \langle Rep(p'), p'\rangle_n\}$, where a belief system T is coherent just in case the size n multiset O consisting of the propositional contents of T is coherent.[4] Evidential independence on Olsson's account is a condition governing witness reports rather than the propositional contents of those reports, since Olsson is interested in the case where n independent witnesses give identical reports.

[4] Olsson states that O is a tuple, rather than a multiset, but order is not important for his proposal. Multiplicity is.

Shogenji (2007) presents a simplified witness model that removes the distinction between evidence reports and the contents of those reports, and removes the logical restriction that identifies the confirmable hypothesis with the conjunction of the evidence contents. EI on his account holds between the evidence contents $A_1, \ldots, A_n$, given a logically unrelated hypothesis, h.

Shogenji's generalized EI-witness model nevertheless preserves the basic distinction between reports and the contents of those reports, which is the key idea underpinning the witness-report strategy. By incorporating evidential independence directly into a model of evidence, Shogenji shows that

> ... under the condition of evidential independence, the degree of coherence is simply a function of the individual strengths of the pieces of evidence. Thus, although there is a sense in which coherence is truth-conducive... the lateral relation, such as coherence, has no independent role to play in the confirmation of the hypothesis (Shogenji 2007, 371).

So, given that the point behind Bonjour's cognitively spontaneous beliefs was to generate coherentist justification without depending upon the individual strengths of each belief, Shogenji's result appears to be bad news for the coherence theory of justification.

Perhaps. But Bonjour's outline of the coherence theory is a rough one, and nowhere does he explicitly endorse EI or the underlying constraints of EI-witness models. Furthermore, it is misleading to draw a general conclusion about the relationship between deviation from independence and confirmation on the basis of Shogenji's results. Shogenji's writes that under EI 'the conditional probability of the hypothesis, $Pr(h|A_1 \cap \ldots A_n)$, is a strictly increasing function of $Pr(h|A_i)$ for each $i = 1, \ldots, n$' (Shogenji 2007, 366, my notation), which we can see from Proposition 1. But this is not true in general, since the second position of the conditional probability function $\Pr(h|A_1 \cap \ldots \cap A_n)$ is *not* a strictly monotone function on the size of n in $\Pr(h|A_i)$, for $1 \leq i \leq n$. The conditional probability function $\Pr(\cdot|\cdot)$ itself is monotone, just as all probability functions are monotone: if $A \subseteq B$ then $\Pr(A|\cdot) \leq \Pr(B|\cdot)$; If A is a smaller set of possibilities than B, then the probability of A must be less than (or equal to) the probability of B. But $\Pr(\cdot|A_i)$ *in general* is not a strictly increasing function of the size of the set of possibilities $\{A_i\}$ for $1 \leq i \leq n$, since if $A \not\equiv B$ then $\Pr(\cdot|A)$ may be greater than, less than, or equal to $Pr(\cdot|A \cap B)$. Conditioning on a smaller set of possibilities may either increase or decrease the probability of the conditioning event.

It is the restriction to strictly increasing conditional measures that is driving Shogenji's result, not evidential independence *per se*. So, we might think that a clearer picture of Shogenji's observation would be given by replacing EI by a weaker monotonicity condition.

Monotone Evidence (ME): A and B are monotone evidence for h if and only if $\Pr(A|B \cap h) \geq \Pr(A|h)$ and $\Pr(A|B \cap \overline{h}) \leq \Pr(A|\overline{h})$.

Adopting ME rather than IE would seem to increase the scope of Shogenji's result.

Even so, Bonjour nowhere endorses anything like either ME or EI. Even if we assume that there is some static notion of coherence behind Bonjour's claims about cognitively spontaneous beliefs, there are a variety of ways in which those beliefs may be understood to be independent from one another. One candidate is the *error assumption* mentioned in section 2.1. Rather than start with a heavy-handed witness-report representation of cognitively spontaneous beliefs that gives you independence *and* dubious structural constraints on the class of models to evaluate, perhaps a better approach would be to investigate the relationship between probabilistic dependence and confirmation where much weaker constraints are maintained, like the observational independence assumption. Assuming observational independence leaves open how to describe the mechanism by which the independence of each observation is generated. One is then left free to consider whether there is an informative relationship between correlation and confirmation after all.

3 Focused Correlation

There is no direct relationship between correlation and confirmation but there is an indirect one: the degree of confirmation of h by both A and B combined is determined by the product of the degree of confirmation of h by A, the degree of confirmation of h by B, and the *ratio* of the correlation of the evidence set $\{A, B\}$ conditioned on h to the correlation of A and B. This point is also made by Myrvold (Myrvold 1996, 663). The relationship between these three factors, expressed by

$$\frac{\Pr(h|A \cap B)}{\Pr(h)} = \frac{\Pr(h|A)}{\Pr(h)} \times \frac{\Pr(h|B)}{\Pr(h)} \times \frac{S(A, B|h)}{S(A, B)}, \tag{5}$$

is proved in the appendix.

Equation (5) suggests a new correlation measure, which is relativized to a particular hypothesis of interest, called *focused correlation*:

$$For_h(A, B) := \frac{S(A, B|h)}{S(A, B)} = \frac{\Pr(h|A \cap B)}{\Pr(h)} \times \frac{\Pr(h)}{\Pr(h|A)} \times \frac{\Pr(h)}{\Pr(h|B)}. \tag{6}$$

The focused correlation of A and B relative to a hypothesis h, $For_h(A, B)$, tells us what impact there is on the confirmation of h, if any at all, from combining A and B.

When we restrict ourselves to cases in which each piece of evidence confirms h, i.e., when $\Pr(h|A) > \Pr(h)$ and $\Pr(h|B) > \Pr(h)$, then values of $For_h(A, B)$ greater than 1 tell us that the evidence set $\{A, B\}$ offers more confirmation to h than A and B alone. Likewise, when $\Pr(h|A) < \Pr(h)$ and $\Pr(h|B) < \Pr(h)$, then values of $For_h(A, B)$ less than 1 tell us that the evidence set $\{A, B\}$ offers even less confirmation to h than A and B alone.

When $\Pr(h|A) > \Pr(h)$ and $\Pr(h|B) > \Pr(h)$, then values for $For_h(A, B)$ less than 1 tell us that the evidence set $\{A, B\}$ does *not* offer more confirmation to h than A and B alone. Even so, the evidence set $\{A, B\}$ may offer confirming evidence for h. The point is that combining A and B won't gain an advantage. Similarly, when $\Pr(h|A) < \Pr(h)$ and $\Pr(h|B) < \Pr(h)$, then values for $For_h(A, B)$ greater than 1 tell us that the evidence set $\{A, B\}$ does not offer less confirmation to h than A and B alone. The evidence set $\{A, B\}$ may still offer disconfirming evidence for h, but the combination will not be a greater disadvantage so long as $For_h(A, B) \neq 0$. What these cases illustrate is that combining confirming (disconfirming) evidence whose focused correlation is negative (positive) will not improve (worsen) the confirmation of h, respectively. The interpretation of these four cases is encapsulated in Figure 1.

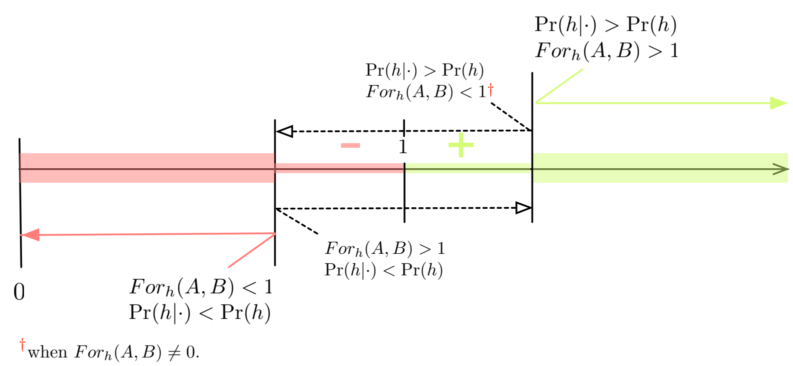

Fig. 1. Interpreting positive and negative values for For_h on incrementally confirming and disconfirming evidence.

The lower limit of $For_h(A, B)$ is 0, which means that the evidence set $\{A, B\}$ offers no information about h. We stipulate that if $For_h(A, B) = \frac{0}{0}$, then $For_h(A, B) = 0$. When $For_h(A, B) = 1$, there is no difference in information about h between conditioning on the information set $\{A, B\}$ and conditioning on A and B separately. Notice that there are cases in which $For_h(A, B) = \frac{0}{0}$ but where there is a difference in information about h between conditioning on the evidence set versus conditioning on focal pieces of evidence. Example 3 is an illustration.

When $\Pr(h|A) = \Pr(h)$ and $\Pr(h|B) = \Pr(h)$, then $For_h(A, B)$ may be positive, negative, or neutral. This is to say that when both A is independent of h, and B is independent of h, then combining A and B may increase the confirmation of h, decrease the confirmation of h, or have no impact on h.

Finally, we may generalize (6) to a focused *dfi*-measure for evidence sets of $1 < n$ distinct, finite variables:

$$For_h(A_1, \ldots, A_n) := \frac{S(A_1, \ldots A_n|h)}{S(A_n, \ldots, A_n)}. \tag{7}$$

Application of focused correlation to the four examples illustrates how this measure may be interpreted as an indicator function to identify when combining confirming pieces of evidence offers more (or less) confirmation to a hypothesis. See the appendix for details.

Observation 1. *Example 1 presents a case where the evidence set* $\{A_1, A_2\}$ *is* negatively correlated*, but the degree of confirmation of* h_1 *on* $\{A_1, A_2\}$ *is* greater than *the degree of confirmation of* h_1 *on the event* A_1*, and greater than the degree of confirmation on the event* A_2*. Note, however, that* A_2 *provides more information about* h_1 *than* A_1 *does. A consequence of this asymmetry is that learning* A_2 *either before or after learning* A_1 *provides more information about* h_1 *than learning* A_1 *either before or after learning* A_2*. Nevertheless, the evidence set* $\{A_1, A_2\}$ *offers positive confirmation for* h_1*, more so than either* A_1 *or* A_2 *alone. The* $For_{h_1}(A_1, A_2)$ *is (approximately)* 1.20.

Observation 2. *Example 2 presents a case where the evidence set* $\{A_3, A_4\}$ *is* positively correlated*, and the degree of confirmation of* h_2 *on* $\{A_3, A_4\}$ *is* greater than *the degree of confirmation of* h_2 *on each event,* A_3 *and* A_4*, individually. The* $For_{h_2}(A_3, A_4)$ *is* 1.6.

Observation 3. *Example 3 presents a case where the evidence set* $\{A_5, A_6\}$ *is* negatively correlated*, and the degree of confirmation of* h_3 *is* less than *the degree of confirmation of* h_3 *on each event,* A_5 *and* A_6*, individually. No new information is learned from* A_5 *after observing* A_6*, nor by* A_6 *after observing* A_5*. In fact, no new information* can *be learned since* A_5 *and* A_6 *are mutually exclusive events. The* $For_{h_3}(A_5, A_6)$ *is* 0.

Observation 4. *Example 4 presents a case where the evidence set* $\{A_7, A_8\}$ *is* positively correlated*, but the degree of confirmation of* h_4 *is* less than *the degree of confirmation of* h_4 *on each event,* A_7 *and* A_8*, individually. The* $For_{h_4}(A_7, A_8)$ *is* 0.4.

Although not intended to be an explication of coherentist justification, focused correlation nevertheless has four attractive features worth noting. First, focused correlation specifies a clear relationship between correlation and confirmation, one that does not rely upon strong conditional independence assumptions. Second, the measure does not place restrictions on the logical relationships between the evidence and the hypothesis. Third, focused correlation incorporates the basic structural features necessary to link confirmation to correlation/divergence from independence. Finally, although the measure is commutative, i.e., $For_h(A, B) = For_h(B, A)$, focused correlation nevertheless can reveal important asymmetries in the incremental confirmation of h on pieces of evidence. There is often a difference between learning A before learning B and learning B before learning A, and focused correlation can be used to exploit such asymmetries in information content. Let's now examine each of these points in some detail.

Regarding the role of independence assumptions, note that none of the first four examples satisfies EI, yet the measure indicates when combining evidence increases the confirmation of the hypothesis. We of course may add constraints such as EI without affecting the behavior of focused correlation, since this is but a restricted case of Equation (5). The next example gives an illustration.

Example 5. An urn contains the billiard balls numbered 1, 2, 14, 15, and the cue ball. Consider the hypothesis h_5 and two pieces of evidence, A_9 and A_{10}:

h_5: The drawn ball is an odd stripe.
A_9: The drawn ball is not solid.
A_{10}: The drawn ball is not even.

Note the following values: $\Pr(h_5) = 0.2$, $\Pr(A_9) = \Pr(A_{10}) = 0.6$, $\Pr(h_5|A_9) = \Pr(h_5|A_{10}) = \frac{1}{3}$, and $S(A_9, A_{10}) \approx 1.11$. So, the evidence set $\{A_9, A_{10}\}$ is positively correlated. Note also that the condition of *evidential independence* is satisfied:

$$\Pr(A_9 \cap A_{10}|h_5) = \Pr(A_9|h_5) \times \Pr(A_{10}|h_5) = 1,$$

and

$$\Pr(A_9 \cap A_{10}|\overline{h}_5) = \Pr(A_9|\overline{h}_5) \times \Pr(A_{10}|\overline{h}_5) = 0.25.$$

However, the set $\{A_9, A_{10}\}$ neither confirms h_9 more than A_9, nor more than A_{10}, i.e.,

$$\frac{\Pr(h_5|A_9)}{\Pr(h_5)} \approx 1.665 \approx \frac{\Pr(h_5|A_{10})}{\Pr(h_5)},$$

but

$$\frac{\Pr(h_5|A_9 \cap A_{10})}{\Pr(h_5|A_{10})} \approx 1.50 \approx \frac{\Pr(h_5|A_9 \cap A_{10})}{\Pr(h_5|A_{10})}.$$

Hence, a positively correlated evidence set $\{A_9, A_{10}\}$ offers less confirmation for h_5 than each piece of evidence alone.

This result is captured by the measure $For_{h_5}(A_9, A_{10})$. Since $S(A_9 \cap A_{10}|h_5) = 1$, therefore

$$For_{h_5}(A_9, A_{10}) = \frac{S(A_9 \cap A_{10}|h_5)}{S(A_9 \cap A_{10})} \approx 0.901.$$

□

The second attractive feature of focused correlation is that it does not place restrictions on the logical relationships between the hypothesis and the evidence set $\{A, B\}$. On Olsson's model and Bovens and Hartmann's model the event h is assumed to be equivalent to the joint event $A \cap B$. Like evidential independence, this assumption is a strong structural constraint that is motivated by representing cognitively spontaneous beliefs in terms of an EI-witness model. Focused correlation does not place this constraint on the logical form of the hypothesis.

The third benefit is that the structure of focused correlation builds in the fact that the relationship between correlation and confirmation is expressed by a specific relationship between a particular correlated evidence set and a particular hypothesis. It makes no more sense to talk about the generic impact that a correlated evidence set has on confirmation without specifying a hypothesis than it does to talk of *between* as a 2- rather than a 3-place relation. The structure of focused correlation makes explicit the basic parameters that are necessary to express the relationship between a correlated evidence set and the confirmation of a specific hypothesis of interest: values for conditional and unconditional forms of the measure $S(\cdot,\cdot)$, a prior distribution for h, and the contribution of confirmational strength that the combination of evidence has over each piece of focal evidence. Note the special case when h is replaced by the evidence set $\{A, B\}$,

$$For_{A\cap B}(A,B) = \frac{\Pr(A)\times\Pr(B)}{\Pr(A|A\cap B)\times\Pr(B|A\cap B)\times\Pr(A\cap B)}, \tag{8}$$

which can be generalized to evidence sets of size n:

$$For_{A_1,\cap\ldots\cap A_n}(A_1,\ldots,A_n) = \frac{\prod_{i=1}^{n}\Pr(A_i)}{\prod_{i=1}^{n}\Pr(A_i|A_1\cap\ldots\cap A_n)\times\Pr(A_1\cap\ldots\cap A_n)}. \tag{9}$$

While the first two features of focused correlation highlight the generality of the measure, the third feature highlights an important restriction, i.e., in order to assess the impact of correlated evidence on the confirmation of a hypothesis it is necessary to select a particular hypothesis. These constraints are captured by the structure of the measure, and equations (8) and (9) simply apply the result to instances where one is interested in the evidence set itself. This last detail leads to our final point, which concerns the tracking of information gain as an agent learns new evidence. $For_h(A,B)$ is commutative because it represents a static evaluation of the factors contributing to confirmation that are embedded in the particular distribution underlying $For_h(A,B)$. Nevertheless, we might be interested in different paths through the evidence that an inquiry may follow. And here the order in which evidence is learned may be very important.

To illustrate consider again Example 1. This example highlights an asymmetry in the information that each focal piece of evidence reveals about h. In addition to knowing the overall impact that the evidence set $\{A_1, A_2\}$ has on h, we might also be interested in exploiting our knowledge that observing A_2 is more informative than observing A_1. The advantage of this ability to rank information impact becomes more apparent for larger evidence sets, for in such cases we may in effect rank evidence by impact on an hypothesis, from greatest impact to least. Furthermore, we may exploit ordered evidence as a method for efficiently moving from a coarse to precise value for h. The measure For_h alone does not yield this information, but calculating values necessary to apply focused correlation to a problem does yield this information.

4 Conclusion

The measure of *focused correlation* explicates the relationship between correlation (deviation from independence) and incremental confirmation. There are several insights this result offers, but the main point is that it resolves a formal question underlying 'truth-conduciveness' arguments found within Bayesian coherentism. In so far as the question raised by probabilistic theories of coherence is whether there is an informative relationship between deviation from independence measures and incremental confirmation, the answer is Yes. How independence of observations is construed is crucial for EI-witness model results, and for the results obtained here. The key to the results here is to assume that observational independence holds without specifying a particular mechanism that generates those independencies. So, in so far as Bayesian coherence theorists are interested in the relationship between correlation and confirmation irrespective of particular mechanisms that guarantee that observations are independent, focused correlation provides an account.

Focused correlation is a static *dfi*-measure, and there are several precautions and caveats that attend drawing inferences from correlation measures on statistical data that are generally ignored by Bayesian epistemologists. Even so, there is an interesting relationship between correlation and confirmation that, with due care, may be exploited for informative inference.

Funding: The Leverhulme Trust; Portuguese Fundação para Ciência e a Tecnologia (SFRH/BPD-13699-2003).

Acknowledgements: Thanks to Marco Castellani, Stephen Fogdall, Tomoji Shogenji, Phiniki Stouppa, Jon Williamson, and the audience at FEW 2008, Madison. Special thanks to Peter Brössel for spotting an error in one of my proofs.

5 Appendix

Proof (Equation (5)).

$$
\begin{aligned}
&\text{Show :}\\
\frac{\Pr(h|A\cap B)}{\Pr(h)} &= \frac{\Pr(h|A)}{\Pr(h)} \times \frac{\Pr(h|B)}{\Pr(h)} \times \frac{S(A,B|h)}{S(A,B)}.\\
&\text{Let :}\\
\alpha &= \frac{S(A,B|h)}{S(A,B)}\\
&= \frac{\frac{\Pr(A\cap B|h)}{\Pr(A|h)\times\Pr(B|h)}}{\frac{\Pr(A\cap B)}{\Pr(A)\times\Pr(B)}}\\
&= \frac{\Pr(A\cap B|h)}{\Pr(A|h)\times\Pr(B|h)} \times \frac{\Pr(A)\times\Pr(B)}{\Pr(A\cap B)}.\\
\beta &= \frac{(h|A\cap B)}{\Pr(h)}\\
&= \frac{\Pr(A\cap B|h)\Pr(h)}{\Pr(A\cap B)} \times \frac{1}{\Pr(h)}\\
&= \frac{\Pr(A\cap B|h)}{\Pr(A\cap B)}.\\
\gamma &= \frac{\Pr(h|A)}{\Pr(h)} \times \frac{\Pr(h|B)}{\Pr(h)}\\
&= \frac{\Pr(A|h)}{\Pr(A)} \times \frac{\Pr(B|h)}{\Pr(B)}.\\
&\text{So,}\\
\alpha &= \frac{\beta}{\gamma}\\
\frac{\Pr(A\cap B|h)}{\Pr(A\cap B)} \times \frac{\Pr(A)}{\Pr(A|h)} \times \frac{\Pr(B)}{\Pr(B|h)} &= \frac{\Pr(A\cap B|h)}{\Pr(A\cap B)} \times \frac{\Pr(A)}{\Pr(A|h)} \times \frac{\Pr(B)}{\Pr(B|h)}.
\end{aligned}
$$

□

Example 1:

$$
\begin{aligned}
S(A_1, A_2) &= \frac{\Pr(A_1 \cap A_2)}{\Pr(A_1)\Pr(A_2)} \\
&= \frac{0.2}{0.24} \\
&\approx 0.833. \\
\Pr(h_1|A_1 \cap A_2) &= \frac{\Pr(A_1 \cap A_2|h_1) \times \Pr(h_1)}{\Pr(A_1 \cap A_2|h_1) \times \Pr(h_1) + \Pr(A_1 \cap A_2|\overline{h}_1) \times \Pr(\overline{h}_1)} \\
&= \frac{1 \times 0.2}{(1 \times 0.2) + (0 \times 0.8)} \\
&= 1. \\
\text{So,} & \\
\frac{\Pr(h_1|A_1 \cap A_2)}{\Pr(h_1|A_1)} &= \frac{1}{\frac{1}{3}} \\
&= 3, \\
\text{and} & \\
\frac{\Pr(h_1|A_1 \cap A_2)}{\Pr(h_1|A_2)} &= \frac{1}{0.5} \\
&= 2 \\
\text{But,} & \\
\frac{\Pr(h_1|A_1)}{\Pr(h_1)} &= \frac{\frac{1}{3}}{0.2} \\
&\approx 1.667, \\
\text{and} & \\
\frac{\Pr(h_1|A_2)}{\Pr(h_1)} &= \frac{0.5}{0.2} \\
&= 2.5.
\end{aligned}
$$

□

Example 2:

$$
\begin{aligned}
S(A_3, A_4) &= \frac{\Pr(A_3 \cap A_4)}{\Pr(A_3)\Pr(A_4)} \\
&= \frac{0.2}{0.16} \\
&= 1.25. \\
\Pr(h_2|A_3 \cap A_4) &= \frac{\Pr(A_3 \cap A_4|h_2) \times \Pr(h_2)}{\Pr(A_3 \cap A_4|h_2) \times \Pr(h_2) + \Pr(A_3 \cap A_4|\overline{h}_2) \times \Pr(\overline{h}_2)} \\
&= \frac{0.5 \times 0.4}{(0.5 \times 0.4) + (0 \times 0.6)} \\
&= 1. \\
\text{So,} & \\
\frac{\Pr(h_2|A_3 \cap A_4)}{\Pr(h_2|A_3)} &= \frac{1}{0.5} \\
&= 2 = \frac{\Pr(h_2|A_3 \cap A_4)}{\Pr(h_2|A_4)}. \\
\text{But,} & \\
\frac{\Pr(h_2|A_3)}{\Pr(h_2)} &= \frac{0.5}{0.4} \\
&= 1.25, \\
\text{and} & \\
\frac{\Pr(h_2|A_4)}{\Pr(h_2)} &= \frac{0.5}{0.4} \\
&= 1.25.
\end{aligned}
$$

□

Example 3:

$$
\begin{aligned}
S(A_5, A_6) &= \frac{\Pr(A_5 \cap A_6)}{\Pr(A_5)\Pr(A_6)} \\
&= \frac{0}{0.16} \\
&= 0. \\
\Pr(h_3|A_5 \cap A_6) &= 0 \\
\frac{\Pr(h_3|A_5 \cap A_6)}{\Pr(h_3|A_5)} &= \frac{0}{0.5} \\
&= 0 = \frac{\Pr(h_3|A_5 \cap A_6)}{\Pr(h_3|A_6)}. \\
\text{But,} \\
\frac{\Pr(h_3|A_5)}{\Pr(h_3)} &= \frac{0.5}{0.4} \\
&= 1.25, \\
\text{and} \\
\frac{\Pr(h_3|A_6)}{\Pr(h_3)} &= \frac{0.5}{0.4} \\
&= 1.25.
\end{aligned}
$$

□

Example 4:

$$
\begin{aligned}
S(A_7, A_8) &= \frac{\Pr(A_7 \cap A_8)}{\Pr(A_7)\Pr(A_8)} \\
&= \frac{0.2}{0.08} \\
&= 2.5. \\
\Pr(h_4|A_7 \cap A_8) &= \frac{\Pr(A_7 \cap A_8|h_4) \times \Pr(h_4)}{\Pr(A_7 \cap A_8|h_4) \times \Pr(h_4) + \Pr(A_7 \cap A_8|\overline{h}_4) \times \Pr(\overline{h}_4)} \\
&= \frac{0.5 \times 0.4}{(0.5 \times 0.4) + (0 \times 0.6)} \\
&= 1. \\
\text{So,} \\
\frac{\Pr(h_4|A_7 \cap A_8)}{\Pr(h_4|A_7)} &= \frac{1}{1} = \frac{\Pr(h_4|A_7 \cap A_8)}{\Pr(h_4|A_8)} \\
\text{But,} \\
\frac{\Pr(h_4|A_7)}{\Pr(h_4)} &= \frac{1}{0.4} \\
&= 2.5, \\
\text{and} \\
\frac{\Pr(h_4|A_6)}{\Pr(h_4)} &= \frac{1}{0.4} \\
&= 2.5.
\end{aligned}
$$

□

Example 5:

$$\begin{aligned}
S(A_9, A_{10}) &= \frac{\Pr(A_9 \cap A_{10})}{\Pr(A_9) \times \Pr(A_{10})} \\
&= \frac{0.6 \times \frac{2}{3}}{0.6 \times 0.6} \\
&\approx 1.111. \\
S(A_9, A_{10}|h_5) &= \frac{\Pr(A \cap B|h)}{\Pr(A|h_5) \times \Pr(B|h_5)} \\
&= \frac{1}{1} \\
&= 1.
\end{aligned}$$

Evidential independence :

$$\begin{aligned}
\Pr(A_9 \cap A_{10}|h_5) &= \Pr(A_9|h_5) \times \Pr(A_{10}|h_5) \\
&= 1 \\
\Pr(A_9 \cap A_{10}|\overline{h}_5) &= \Pr(A_9|\overline{h}_5) \times \Pr(A_{10}|\overline{h}_5) \\
&= \frac{1}{2} \times \frac{1}{2} \\
&= 0.25.
\end{aligned}$$

Shogenji's :

$$\begin{aligned}
S(A_9, A_{10}) - 1 &= \frac{(\Pr(h_5|A_9) - \Pr(h_5))(\Pr(\Pr(h_9|A_{10} - \Pr(h_5))}{\Pr(h_5)(1 - \Pr(h_5))} \\
&= \frac{(\frac{1}{3} - 0.2)(\frac{1}{3} - 0.2)}{(0.2)(0.8)} \\
&\approx 0.1111.
\end{aligned}$$

□

Proof (Observation 1). From Example 1, $S(A_1, A_2) \approx 0.833$, and

$$\begin{aligned}
S(A_1, A_2|h_1) &= \frac{\Pr(A_1 \cap A_2|h_1)}{\Pr(A_1|h_1) \times \Pr(A_2|h_1)} \\
&= \frac{1}{1 \times 1} \\
&= 1.
\end{aligned}$$

So,

$$\begin{aligned}
For_{h_1}(A_1, A_2) &= \frac{S(A_1, A_2|h_1)}{S(A_1, A_2)} \\
&\approx \frac{1}{0.833} \\
&\approx 1.20
\end{aligned}$$

□

Proof (Observation 2). From Example 2, $S(A_3, A_4) = 1.25$, and

$$
\begin{aligned}
S(A_3, A_4|h_2) &= \frac{\Pr(A_3 \cap A_4|h_2)}{\Pr(A_3|h_2) \times \Pr(A_4|h_2)} \\
&= \frac{0.5}{0.5 \times 0.5} \\
&= 2. \\
\text{So,} & \\
For_{h_2}(A_3, A_4) &= \frac{S(A_3, A_4|h_2)}{S(A_3, A_4)} \\
&= \frac{2}{1.25} \\
&= 1.6.
\end{aligned}
$$

□

Proof (Observation 3). From Example 3, $S(A_5, A_6) = 0$, and

$$
\begin{aligned}
S(A_5, A_6|h_3) &= \frac{\Pr(A_5 \cap A_6|h_3)}{\Pr(A_5|h_3) \times \Pr(A_6|h_3)} \\
&= \frac{0}{0.5 \times 0.5} \\
&= 0. \\
\text{So,} & \\
For_{h_3}(A_5, A_6) &= \frac{S(A_5, A_6|h_3)}{S(A_5, A_6)} \\
&= \frac{0}{0} \\
&= 0, \text{ by definition.}
\end{aligned}
$$

□

Proof (Observation 4). From Example 4, $S(A_7, A_8) = 2.5$, and

$$\begin{aligned} S(A_7, A_8|h_4) &= \frac{\Pr(A_7 \cap A_8|h_4)}{\Pr(A_7|h_4) \times \Pr(A_8|h_4)} \\ &= \frac{0.5}{0.5 \times 1} \\ &= 1. \end{aligned}$$

So,

$$\begin{aligned} For_{h_4}(A_7, A_8) &= \frac{S(A_7, A_8|h_4)}{S(A_7, A_8)} \\ &= \frac{1}{2.5} \\ &= 0.4. \end{aligned}$$

□

Proof (Proposition 1). Show that

(i): $\Pr(A|B \cap h) = \Pr(A|h) \Leftrightarrow \Pr(A \cap B|h) = \Pr(A|h) \times \Pr(B|h)$,

(ii): $\Pr(A|B \cap \overline{h}) = \Pr(A|\overline{h}) \Leftrightarrow \Pr(A \cap B|\overline{h}) = \Pr(A|\overline{h}) \times \Pr(B|\overline{h})$, and

(iii): $\Pr(A|B \cap h) = \Pr(A|h) \Leftrightarrow$

$$\Pr(h|A \cap B) = \frac{\Pr(A|h)\Pr(B|h)\Pr(h)}{\Pr(A|h)\Pr(B|h)\Pr(h) + \Pr(A|\overline{h})\Pr(B|\overline{h})\Pr(\overline{h})}$$

.

(i) : $\Pr(A|B \cap h) = \Pr(A|h) \Leftrightarrow$

$$\begin{aligned} \frac{\Pr(A \cap B \cap h)}{\Pr(B \cap h)} = \frac{\Pr(A \cap h)}{\Pr(h)} &\Leftrightarrow \frac{\Pr(A \cap B \cap h)}{\Pr(B \cap h)} = \frac{\Pr(A \cap h)}{\Pr(h)} \\ &\Leftrightarrow \frac{\Pr(A \cap B \cap h)}{\Pr(h)} \times \frac{\Pr(h)}{\Pr(B \cap h)} = \frac{\Pr(A \cap h)}{\Pr(h)} \\ &\Leftrightarrow \frac{\frac{\Pr(A \cap B \cap h)}{\Pr(h)}}{\frac{\Pr(A \cap h)}{\Pr(h)}} = \frac{\Pr(A \cap h)}{\Pr(h)} \\ &\Leftrightarrow \frac{\Pr(A \cap B \cap h)}{\Pr(h)} = \frac{\Pr(A \cap h)}{\Pr(h)} \times \frac{\Pr(B \cap h)}{\Pr(h)} \\ &\Leftrightarrow \Pr(A \cap B|h) = \Pr(A|h) \times \Pr(B|h). \end{aligned}$$

(ii) : Substitute $\overline{h}/h$ in (i).

(iii) Iff (i) and (ii), then

$$\begin{aligned} \Pr(h|A \cap B) &= \frac{\Pr(A \cap B|h)\Pr(h)}{\Pr(A \cap B|h)\Pr(h) + \Pr(A \cap B|\overline{h})\Pr(\overline{h})} \\ &= \frac{\Pr(A|h)\Pr(B|h)\Pr(h)}{\Pr(A|h)\Pr(B|h)\Pr(h) + \Pr(A|\overline{h})\Pr(B|\overline{h})\Pr(\overline{h})}. \end{aligned}$$ □

Proof (Equation (8)).

$$\text{Show :}$$
$$For_{A\cap B}(A,B) = \frac{S(A,B|A\cap B)}{S(A,B)} = \frac{\Pr(A\cap B|A\cap B)}{\Pr(A\cap B)} \times \frac{\Pr(A)}{\Pr(A|A\cap B)} \times \frac{\Pr(B)}{\Pr(B|A\cap B)}$$
$$= \frac{\Pr(A)\times\Pr(B)}{\Pr(A\cap B)\times\Pr(A|A\cap B)\times\Pr(B|A\cap B)}.$$

□

Proof (Equation (9)).

$$\text{Show :}$$
$$For_{A_1\cap\ldots\cap A_n}(A_1,\ldots,A_n) = \frac{S(A_1,\ldots,A_n|A_1\cap\ldots\cap A_n)}{S(A_1,\ldots,A_n)}$$
$$\text{Where :}$$
$$S(A_1,\ldots,A_n|A_1\cap\ldots\cap A_n) = \frac{\Pr(A_1\cap\ldots\cap A_n|A_1\cap\ldots\cap A_n)}{\prod_{i=1}^n \Pr(A_i|A_1\cap\ldots\cap A_n)}$$
$$= \frac{1}{\prod_{i=1}^n \Pr(A_i|A_1\cap\ldots\cap A_n)},$$
$$\text{and :}$$
$$S(A_1,\ldots,A_n) = \frac{\Pr(A_1\cap\ldots\cap A_n)}{\prod_{i=1}^n \Pr(A_i)}.$$
$$\text{Hence :}$$
$$= \frac{1}{\prod_{i=1}^n \Pr(A_i|A_1\cap\ldots\cap A_n)} \times \frac{\prod_{i=1}^n \Pr(A_i)}{\Pr(A_1\cap\ldots\cap A_n)}$$
$$= \frac{\prod_{i=1}^n \Pr(A_i)}{\prod_{i=1}^n \Pr(A_i|A_1\cap\ldots\cap A_n)\times\Pr(A_1\cap\ldots\cap A_n)}.$$

□

Bibliography

Bolstad, W. (2004). *Introduction to Bayesian Statistics.* Wiley and Sons, Hoboken, NJ.

Bonjour, L. (1985). *The Structure of Empirical Knowledge.* Harvard University Press, Cambridge, MA.

Bovens, L. and Hartmann, S. (2003). *Bayesian Epistemology.* Oxford University Press, Oxford.

Cohen, L. J. (1977). *The Probable and the Provable.* Clarendon Press, Oxford.

Earman, J. (2000). *Hume's Abject Failure.* Oxford University Press, New York.

Howson, C. and Urbach, P. (1989). *Scientific Reasoning: The Bayesian Approach.* Open Court, La Salle, IL.

Klein, P. and Warfield, T. (1994). What price coherence? *Analysis*, 54(3):129–32.

Klein, P. and Warfield, T. (1996). No help for the coherentist. *Analysis*, 56(2):118–21.

Myrvold, W. (1996). Bayesianism and diverse evidence: A reply to Andrew Wayne. *Philosophy of Science*, 63:661–665.

Olsson, E. (2005). *Against Coherence: Truth, Probability and Justification.* Oxford University Press, Oxford.

Shogenji, T. (1999). Is coherence truth conducive? *Analysis*, 59:338–45.

Shogenji, T. (2007). Why does coherence appear truth-conducive? *Synthese*, 157:361–372. Manuscript for 2006 APA Eastern Division Meeting.

Wayne, A. (1995). Bayesianism and diverse evidence. *Philosophy of Science*, 62(1):111–121.