A Probabilistic Theory of the Coherence of an Information Set

Stephan Hartmann and Luc Bovens

1. The Riddle

Bonjour (1985: 101 and 1999: 124) and other coherence theorists of justification before him (e.g. Ewing, 1934: 246) have complained that we do not have a satisfactory analysis of the notion of coherence. The problem with existing accounts of coherence is that they try to bring precision to our intuitive notion of coherence independently of the particular role that it is meant to play within the coherence theory of justification (e.g Lewis, 1946: 338). This is a mistake: it does not make any sense to ask what precisely makes for a more coherent information set independently of the particular role that coherence is supposed to play within the context in question. What is this context and what is this role? The coherence theory of justification rides on a particular common sense intuition: when we gather information from less than fully reliable sources, then the more coherent the story that materializes is, the more confident we may be, *ceteris paribus*. Within the context of information gathering from certain types of sources, coherence is a property of stories which plays a confidence boosting role. But what features should the information sources have, so that the coherence of the information set is indeed a determinant of our degree of confidence in question? And what goes into the *ceteris paribus* clause? In other words, what other factors affect our confidence in the information set in question?

Suppose that one receives information from independent and relatively unreliable sources. What determines our degree of confidence in the information set? Intuitively, we can see three distinct factors: (i) How surprising is the information? (ii) How reliable are the sources? (iii) How coherent is the information? Each of these factors has an independent impact on our degree of confidence.

First, suppose that the sources are halfway reliable and the information is halfway coherent. Then certainly our degree of confidence will be greater when the reported information is less rather than more surprising, or in other words, is more rather than less expected. Second, suppose that the information is halfway surprising and is halfway coherent. Then certainly our degree of confidence will be greater when we know our sources to be more like truth-tellers than when we know them to be more like randomizers. Third, consider a scientist who runs two independent tests to determine the locus of a genetic disease on the human genome. In the first case, the tests respectively point to two fairly narrow regions that almost completely overlap in a particular section *s*. In the second case, the tests respectively point to fairly broad regions that overlap in the very same section *s*. Suppose that the tests are halfway reliable and this section is a somewhat surprising locus for the disease. Then certainly the degree of confidence that the locus of the disease is in this region is greater in the former case, in which the information is more coherent, than in the latter case, in which the information is less coherent.

We will construct a model in order to define measures for each of these determinants. It is easy to construct an expectance measure and a reliability measure. As to coherence, the matter is more complex: there does not exist a quantitative coherence measure as such, but we will be able to define a measure that yields a partial coherence ordering over information sets.

2. The Model

Suppose that there are n independent and relatively unreliable sources and each source i informs us of a proposition R_i, for i=1,...,n, so that the information set is $\{\mathrm{R}_1,...,\mathrm{R}_n\}$. For each proposition R_i (in roman script) in the information set, let us define a propositional variable R_i (in italic script) which can take on two values, viz. R_i and $\overline{\mathrm{R}_i}$ (i.e. not-R_i), for i=1,...,n. Let $REPR_i$ be a propositional variable which can take on two values, viz. REPR_i, i.e. after consultation with the proper source, there is a report to the effect that

R_i is the case, and $\overline{REPR_i}$, i.e. after consultation with the proper source, there is no report to the effect that R_i is the case. We construct a joint probability distribution $\boldsymbol{P}$ over $R_1,\ldots, R_n, REPR_1,\ldots, REPR_n$, satisfying the constraint that the sources are independent and relatively unreliable.

We model the independence of the sources by stipulating that $\boldsymbol{P}$ respects the following conditional independences:

(1) $I(\{REPR_i\}, \{R_1, REPR_1,\ldots, R_{i-1}, REPR_{i-1}, R_{i+1}, REPR_{i+1},\ldots, R_n, REPR_n\}|R_i)$ for i=1,...,n

or, in words, $REPR_i$ is probabilistically independent of R_1, $REPR_1,\ldots, R_{i-1}$, $REPR_{i-1}$, R_{i+1}, $REPR_{i+1},\ldots, R_n$, $REPR_n$, given R_i, for i=1,..., n. What this means is that the probability that we will receive a report that R_i given that R_i is the case (or is not the case), is not affected by any additional information about whether any other propositions are the case or whether there is a report to the effect that any other proposition is the case.

We make the simplifying assumption that our relatively unreliable sources are all equally reliable. We specify the following two parameters: $\boldsymbol{P}(REPR_i|R_i) = p$ and $\boldsymbol{P}(REPR_i|\overline{R_i}) = q$ for i=1,...,n. If the information sources would be truth-tellers, then q=0, while if they would be randomizers, then p=q. Since relatively unreliable information sources are more reliable than randomizers, but less reliable than truth-tellers, we impose the following constraint on $\boldsymbol{P}$:

(2) $p > q > 0$

The degree of confidence in the information set is the posterior joint probability of the propositions in the information set after all the reports have come in:

(3) $\boldsymbol{P}^*(R_1,\ldots,R_n) = \boldsymbol{P}(R_1,\ldots,R_n|REPR_1,\ldots, REPR_n)$.

It can be shown[1] that, given the constraints on $\boldsymbol{P}$ in (1) and (2),

(4) $$\boldsymbol{P}^*(R_1,\ldots,R_n) = \frac{a_0}{\sum_{i=0}^{n} a_i x^i},$$

in which the *likelihood ratio* x:= q/p and a_i is the sum of the joint probabilties of all combinations of values of the variables $R_1,\ldots, R_n$ that have i negative values and n-i positive values. For example, for an information triple containing the propositions R_1, R_2, and R_3, $a_2=\boldsymbol{P}(\overline{R_1},\overline{R_2},R_3)+\boldsymbol{P}(\overline{R_1},R_2,\overline{R_3})+\boldsymbol{P}(R_1,\overline{R_2},\overline{R_3})$. Figure 1 contains the probability space which represents a joint probability distribution over the propositional variables R_1, R_2, R_3 and contains the corresponding values for a_i, for i=0,..., 3. Note that $\sum_{i=0}^{n} a_i = 1$. Suppose that the sources are twice as likely to report that R_i is the case, when it is the case, as then, when it is not the case, so that x = 1/2. Then our degree of confidence after we have received the reports from the sources is

(5) $$\boldsymbol{P}^*(R_1,\ldots,R_n) = \frac{.05}{.05\times.5^0+.30\times.5^1+.45\times.5^2+.20\times.5^3} \approx .15$$

a_0=.05

$a_1=3\times.10$ = .30

$a_2=3\times.15$ = .45

a_3=.20

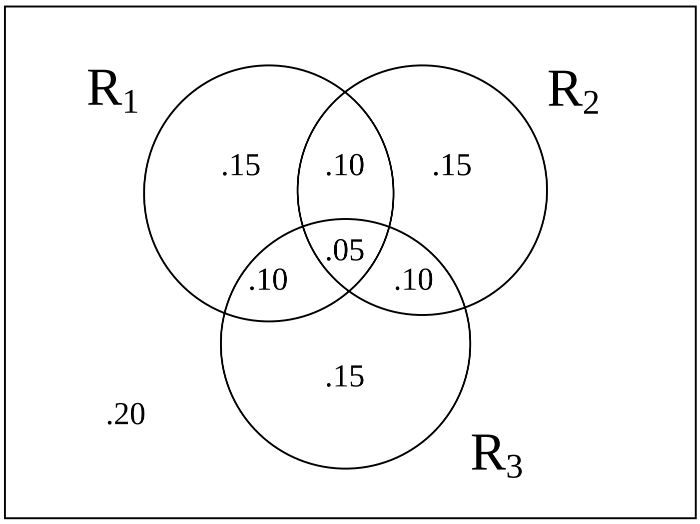

Figure 1

3. Expectance, Reliability and Coherence

We can directly identify the first determinant of the degree of confidence in the information set. Note that $a_0 = \boldsymbol{P}(R_1,...,R_n)$ is the prior joint probability of the propositions in the information set, i.e. the probability before any information was received. This prior probability is lower for more surprising information and higher for less surprising information. Since more surprising information is tantamount to less expected information, let us call this prior probability the *expectance measure*.

We can also directly identify the second determinant, i.e. the reliability of the sources. Note that $\boldsymbol{P^*}(R_1,...,R_n)$ in (4) is a monotonically decreasing function of the likelihood ratio x=q/p. Hence, let us call r:=1-x the *reliability measure*, since $\boldsymbol{P^*}(R_1,...,R_n)$ is a monotonically increasing function of r and the limits of this measure are 0 for sources that are randomizers and 1 for sources that are truth-tellers.[2]

Let us now turn to the third determinant, viz. the coherence of the information set. The coherence of the information set is some function of the series $<a_0,..., a_n>$ that is associated with the joint probability distribution over $R_1,..., R_n$. A maximally coherent information set has the associated series $<a_0, 0, ..., 0, 1-a_0>$: in this case, all items of information $R_1,..., R_n$ are coextensive. If $a_1,...,$ or a_{n-1} exceed 0, then the information set loses this maximal coherence. But it is not clear yet what function of $<a_0,..., a_n>$ determines the coherence of the information set.

Let us first construct a normalized measure. Consider the following analogy: to assess the impact of a training program, we measure the rate of a student's actual performance level after training in the program over the performance level that he would have reached after training in an ideal program, *ceteris paribus*. Similarly, to assess the impact of coherence, we consider the rate of the actual degree of confidence over the degree of confidence that would have obtained had the information set been maximally coherent, *ceteris paribus*. The information set would have been maximally coherent, *ceteris paribus*, if and only if each information source i would have provided precisely the information $\bigcap_{i=1}^{n} R_i$. The idea is simple: in our earlier example of independent tests that identify sections on the human genome that contain the locus of a genetic disease, we consider the counterfactual situation in which both tests would have identified precisely the section *s*. In this case, the tests would have yielded maximally coherent information. So what would our degree of confidence have been, had the information been maximally coherent, *ceteris paribus*? Let ***P*** be the actual joint probability distribution for $R_1,..., R_n$. Construct a joint probability distribution $\boldsymbol{P}^{max}$ with the same expectance measure and the same reliability measure as ***P***, but on $\boldsymbol{P}^{max}$, $R_1,...,$ and R_n are all coextensive and coincide precisely with $\bigcap_{i=1}^{n} R_i$ on ***P***—i.e. $\boldsymbol{P}^{max}(R_i|R_j)=1$ for i,j = 1,..., n and $\boldsymbol{P}^{max}(R_1) = ... = \boldsymbol{P}^{max}(R_n) = \boldsymbol{P}(\bigcap_{i=1}^{n} R_i)$. Then $a_0^{max}= a_0$ and $a_n^{max}= 1-a_0$, so that $a_i^{max}= 0$, for all i=1,..., n-1. $\boldsymbol{P}^{max}$ for the information triple in our earlier example is presented in figure 2. We define $\overline{a_0} := 1-a_0$. It follows directly from (4) that

$$(6) \qquad \boldsymbol{P}^{max}*(R_1,...,R_n)= \frac{a_0}{a_0 + \overline{a_0}x^n} = \frac{.05}{.05 \times .5^0 + .95 \times .5^3} \approx .30 .$$

Hence, for $a_0 \neq 0$, the ratio

$$(7) \qquad c_x(R_1,...,R_n) = \frac{\boldsymbol{P}*(R_1,...,R_n)}{\boldsymbol{P}^{max}*(R_1,...,R_n)} = \frac{a_0 + \overline{a_0}x^n}{\sum_{i=0}^{n} a_i x^i}$$

is a measure of the impact of the coherence of the information set on the degree of confidence in the information set.

a_0=.05

$a_1=3 \times 0 = 0$

$a_2=3 \times 0 = 0$

a_3=.95

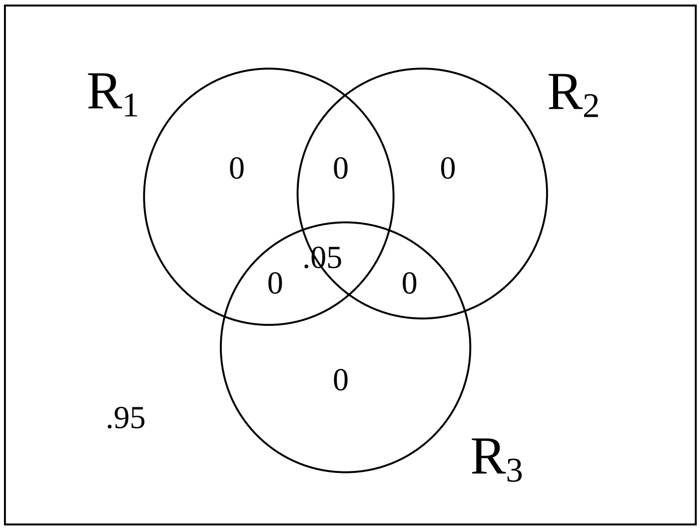

Figure 2

Note that the measure in (7) is contingent on x, and hence on the reliability of the sources. This is unwelcome: our pretheoretical notion of the coherence of an information set has nothing to do with the reliability of the sources that provides us with its content. On the other hand, this pretheoretical notion is an ordinal rather than a cardinal notion. And furthermore, respecting vagueness, we would expect a partial rather than a complete ordering over information sets: for certain, though not for all pairs of information sets, a judgment of relative coherence is in order.

The measure in (7) permits us to construct a partial ordering over information sets which is not contingent on the reliability of the sources. For some pairs of information sets {S,S'}, c_x(S) will always be greater than c_x(S'), no matter what value we choose for x. In this case, S is more coherent than S'. For other pairs of information sets {T,T'}, c_x(T) is greater than the measure for c_x(T') for some values of x and smaller for other values of x. In this case, there is no fact of the matter which is the more coherent set. We will see that this distinction squares with our willingness to make intuitive judgments about which of two information sets is more coherent.

Formally, consider two information sets S={R_1,..., R_m} and S'={R_1',..., R_n'} and let ***P*** be the joint probability distribution for $R_1,..., R_m$ and ***P'*** the joint probability distribution for $R_1',..., R_n'$. We calculate the series

$<a_0,...,a_m>$ for ***P*** and $<a_0',...,a_n'>$ for ***P'*** and construct the following *difference function*:

(8) $f_x(S,S') = c_x(S) - c_x(S')$

$f_x(S,S')$ has the same sign for all values of x ranging over the open interval (0,1) if and only if the measure $c_x(S)$ is always greater than or is always smaller than the measure $c_x(S')$, for any value of x in this interval. We define an ordinal measure for coherence:

(9) For two information sets S and S', $S \succeq S'$ if and only if $f_x(S,S') \geq 0$ for all values of $x \in (0,1)$

for '$\succeq$' denoting the binary relation of *being more coherent than or equally coherent as*, defined over information sets. This relation induces a partial ordering over a set of information sets.

If the information sets S and S' are of equal size, then it is also possible to determine whether there exists a coherence ordering over these sets *directly* from the joint probability distributions ***P*** (with the associated series $<a_0,...,a_n>$) and ***P'*** (with the associated series $<a_0',..., a_n'>$): one needs to evaluate the conditions under which the sign of the difference function is invariable for all values of $x \in (0,1)$. It can be shown that

(10) (i) $a_i' \geq a_i \;\&\; a_0'/a_0 \leq a_i'/a_i,\ \forall\, i=1,...,n-1$, or,
(ii) $a_i' \leq a_i \;\&\; a_0'/a_0 \geq a_i'/a_i,\ \forall\, i=1,...,n-1$

is a sufficient condition for the existence of a coherence ordering over {S, S'} for n>2 and a necessary and sufficient condition for n=2. Furthermore, we can also determine the *direction* of this ordering. It can be shown that

(11) (i) $a_0' \leq a_0 \;\&\; a_i' \geq a_i,\ \forall\, i=1,...,n-1$, or,
(ii) $a_0' \geq a_0 \;\&\; a_i'/a_i \geq a_0'/a_0,\ \forall\, i=1,...,n-1$,

is a sufficient condition for $S \succeq S'$ for n>2 and necessary and sufficient condition for n=2.

4. A Corpse in Tokyo

Does our analysis yield the correct results for some intuitively obvious cases? We will restrict ourselves here to a comparison of two information sets of size 2. Suppose that we are trying to locate a corpse of a murder somewhere in Tokyo. We draw a grid of 100 squares over the map of the city so that it is equally probable that the corpse is hidden in each square. We interview two relatively unreliable sources. In the base case *a*, source 1 reports that the corpse is somewhere in squares 50 to 60 and source 2 reports that the corpse is somewhere in squares 51 to 61. We include this information in the information set S^a. For this information set, $a_0{}^a$=.10 and $a_1{}^a$=.02.

Let us now alter the information from the sources. In the alternate case *b*, source 1 reports squares 20 to 55 and source 2 reports squares 55 to 90. We include this information in S^b. The overlapping area shrinks to $a_0{}^b$=.01 and the non-overlapping area expands to $a_1{}^b$=.70. In the alternate case *c*, source 1 reports squares 20 to 61 and source 2 reports squares 50 to 91. Again we include this information in S^c. The overlapping area expands to $a_0{}^c$=.12 and the non-overlapping area expands to $a_1{}^c$=.60. On condition (11), S^b and S^c are more coherent than S^a. In these alternate scenarios, the information sets are clearly less coherent than the information set in the base case *a*. In alternate case *b*, the overlap is minimal, and in alternate case *c*, the overlap is only slightly greater than in the base case *a*, while in both cases the sources are making a much broader sweep over the grid.

But now consider a pair of cases in which no ordering of the information sets is possible. In case *d*, source 1 reports squares 41 to 60 and source 2 reports squares 51 to 70: $a_0{}^d$=.10 and $a_1{}^d$=.20. In case *e*, source 1 reports squares 26 to 60 and source 2 reports 41 to 75: $a_0{}^e$=.20 and $a_1{}^e$=.30. Is the information set in case *d* more or less coherent than in case *e*? Notice that, on condition (11), S^d is neither more nor less coherent than S^e. This is in line with our intuitions: In cases like these, we are not tempted to make any pronouncements about the relative coherence of the information set. On the one hand, the proportions of the areas of overlap within the total reported areas is greater in case *e* than in case *d*, which seems to favor S^e as the more coherent set, while, on the other hand, the overlap in S^d is more narrow and both sources have provided more accurate information, which seems to favor S^d as the more coherent set: There simply is no fact of the matter which information set is more coherent.

We have reached these results by applying the special conditions in (11) for comparing information pairs. The same results can be obtained by using our general method in (9). Construct the following difference functions and examine the sign of these functions for all values of $x \in (0,1)$:

(12) $$f_x(S^i,S^j) = c_x(S^i) - c_x(S^j) = \frac{a_0^i + \overline{a_0^i}\,x^2}{a_0^i + a_1^i x + a_2^i x^2} - \frac{a_0^j + \overline{a_0^j}\,x^2}{a_0^j + a_1^j x + a_2^j x^2}$$

for $i=a$ and $j=b$; for $i=a$ and $j=c$; for $i=d$ and $j=e$. As we can see in figure 3, the $f_x(S^a,S^b)$ and $f_x(S^a,S^c)$ are positive for all values of $x \in (0,1)$: S^a is more coherent than S^b and S^c. But $f_x(S^d,S^e)$ is positive for some values and negative for other values of $x \in (0,1)$: there is no fact of the matter whether S^d is more or less coherent than S^e.

This is only one illustration of how our criterion in (9) yields results that are in line with our intuitive judgments of coherence. In this case, we compared information sets of size n=2. In subsequent work, we have investigated how our criterion performs for information sets of equal size for n>2. For instance, Bonjour poses the challenge to provide a principled account that the information set {[All Ravens are black], [This bird is a raven], [This bird is black]} is more coherent than {[This chair is brown], [Electrons are negatively charged], [Today is Thursday]} (1985: 96). We specified plausible probability distributions for these information sets and show how our analysis provides results that meet Bonjour's challenge. Furthermore, we have also tested our criterion for information sets of unequal size and again our results were very much in line with our intuitive judgments of coherence.[3]

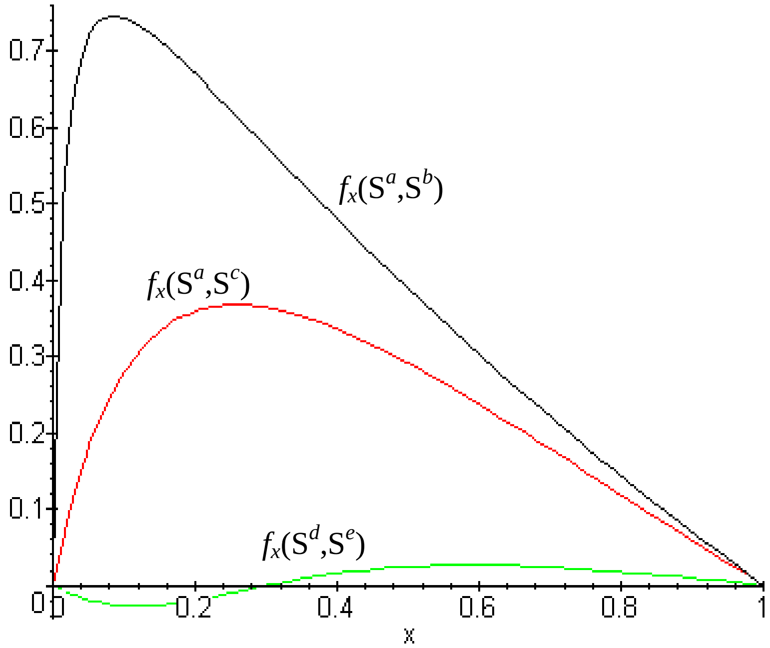

Figure 3

Notes

[1] The proof is straightforward: Apply Bayes Theorem to the right-hand side of (3); simplify on grounds of the conditional independences in (1) and substitute in the parameters p and q as defined in (2); the resulting expression will be well-defined, since, by (2), p>0 and q>0; divide numerator and denominator by p^n; substitute in the parameters x and a_i for i=1,...n as defined underneath.

[2] Bovens and Olsson (2000) investigate under what interpretations of *relatively unreliability* the degree of confidence in the content of the information set is raised by its coherence.

[3] The research was supported by a grant of the National Science Foundation, Science and Technology Studies (SES 00-80580) and grants of the Transcoop Program and of the Feodor Lynen Program of the Alexander von Humboldt Foundation.

References

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

Bonjour, L. (1999): The Dialectic of Foundationalism and Coherentism, in: Greco, J. and Sosa, E. (eds.): *The Blackwell Guide to Epistemology*. Malden [MA]: Blackwell, pp. 117-42.

Bovens, L. and Olsson E.J. (2000): Coherentism, Reliability and Bayesian Networks, in *Mind*, 109, pp. 685-719.

Ewing, A.C. (1934): *Idealism: a Critical Survey*. London: Methuen.

Lewis, C.I. (1946): *An Analysis of Knowledge and Valuation*. LaSalle [ILL]: Open Court.