Bayesian Epistemology

Luc Bovens
and
Stephan Hartmann

CLARENDON PRESS · OXFORD

1

Information

1.1. C. I. LEWIS'S HERITAGE

Laurence BonJour (1985: 97, 147–8) draws our attention to some passages in C. I. Lewis's *An Analysis of Knowledge and Valuation*. Lewis argues that the degree of confidence that we have in n information items (for $n \geq 2$) gathered from independent and partially reliable witnesses[1] is positively affected by their *congruence* (1946: 243–53). The core idea is that the more congruent the information is, i.e. the better it meshes or fits together, the more confident we may be that the information is true. On one side of the continuum there is full congruence, viz. when the witnesses all provide us with precisely the same information. On the other side of the continuum there is complete lack of congruence, viz. when the witnesses provide us with items of information that are mutually exclusive. Between these extremes there are various gradations. Suppose that we are informed by one witness that a particular person drives a Porsche and by another witness that he is a millionaire. This is more congruent information than when we are informed by one witness that the person in question drives a Porsche and by another witness that he is homeless. Lewis (1946: 338) only distinguishes between congruent and non-congruent information sets and proposes a probabilistic criterion to draw this distinction.

In BonJour, 'coherence'[2] has been substituted for 'congruence', because of its role in the coherence theory of justification. We will present a precise and seemingly plausible interpretation of Lewis's claim and will name this interpretation 'Bayesian Coherentism'. Then

[1] Lewis actually talks about 'relatively unreliable witnesses' (1946: 346). Following Olsson (2002*b*: 259), we substitute 'partial reliability' for 'relative unreliability'.

[2] Coherence is a property of information *sets*. At some junctions we also talk about the coherence of information, the coherence of reports, or the coherence of particular propositions for ease of presentation. Nothing hangs on this and each such occurrence can readily be rephrased in terms of information sets.

we will construct a model of independent and partially reliable witnesses to evaluate whether Bayesian Coherentism is defensible.

The results of our analysis are twofold. First, on our interpretation of Lewis, Bayesian Coherentism will turn out to be too strong a thesis. Our analysis will show why the quest for a probabilistic measure that induces a coherence ordering over information sets is in vain. Second, our analysis will suggest a way to salvage certain intuitions that underlie Bayesian Coherentism. This can be achieved if we accept that there cannot exist a coherence *ordering*, but only a coherence *quasi-ordering*, over information sets.

I.2. BAYESIAN COHERENTISM

Suppose that we receive items of information from independent and partially reliable sources, say, observations, witness consultations, experimental tests, etc. Then what determines our degree of confidence that the conjunction of these items of information is true? Consider the following procedure. We are trying to determine the *locus* of the faulty gene on the human genome that is responsible for a particular disease. Before conducting the experiments, there are certain *loci* that we consider to be more likely candidates. We run two tests with independent and partially reliable instruments. Each test identifies an area on the human genome where the faulty gene might be located. It turns out that there is a certain overlap between the indicated areas. It is plausible that the following three factors affect our degree of confidence that *both* tests are providing us with correct data, i.e. that the faulty gene is indeed located somewhere in the overlapping area.

(i) *How expected are the results?* Compare two cases of the above procedure. Suppose that the only difference between the cases is that, given our background knowledge, in one case the overlapping area is initially considered to be a highly expected candidate area, whereas in the other case the overlapping area is initially considered to be a highly unexpected candidate area for the faulty gene. Then clearly, our degree of confidence that the *locus* of the faulty gene is in the overlapping area will be lower in the latter than in the former case.

(ii) *How reliable are the tests?* Again, compare two cases of the above procedure. Suppose that the only difference between the cases is that in one case the tests are highly reliable, whereas in the other case they are highly unreliable. Then clearly, our degree of confidence that the *locus* of the faulty gene is in the overlapping area will be lower in the latter than in the former case.

(iii) *How coherent is the information?* This time suppose that the only difference is that in one case both tests identify precisely the same relatively narrow area, whereas in the other case each test identifies a broad area for possible *locus* with an overlap between these areas that coincides with the relatively narrow area in the first case. Then clearly, our degree of confidence that the *locus* of the faulty gene is in the overlapping area will be lower in the latter than in the former case.

The standard way of expressing these claims is as ceteris paribus claims. When gathering information from independent and partially reliable sources, the following claims seem to hold true. First, the more expected (or equivalently, the less surprising) the information is, the greater our degree of confidence, ceteris paribus. Second, the more reliable the information sources are, the greater our degree of confidence, ceteris paribus. Third, the more coherent the information is, the greater our degree of confidence that the information is true, ceteris paribus.

The third claim is a core claim of Bayesian Coherentism. To make it more precise we introduce the following terminology. Let us assume that we obtain the information items $R_1, \ldots, R_n$ from n independent and partially reliable sources. Then $S = \{R_1, \ldots, R_n\}$ is an information set. Now let $\mathcal{S}$ be a set of such information sets. Then the following is the first tenet of Bayesian Coherentism:

(BC_1) For all information sets $S, S' \in \mathcal{S}$, if S is no less coherent than S', then our degree of confidence that the content of S (i.e. the conjunction of the propositions in S) is true is no less than our degree of confidence that the content of S' is true, ceteris paribus.

What the ceteris paribus clause in (BC_1) indicates is the following. The impact of the coherence of the information set on our degree of confidence satisfies (BC_1), assuming that how expected the information is and how reliable the sources are does not vary from information set to information set. Certainly our degree of confidence in less coherent information is, on occasion, greater than our degree of confidence in more coherent information. This may happen when the less coherent information is more expected or when the corresponding witnesses are more reliable. But a Bayesian Coherentist contends that the coherence of an information set increases our degree of confidence assuming that we keep all other relevant factors, viz. the expectance and the reliability, fixed.

What exactly constitutes the coherence of an information set? A number of proposals have been put forward over the years. Lewis (1946: 338) proposes the following criterion: '*A set of statements, or a set of supposed facts asserted, will be said to be congruent if and only if they are so related that the antecedent probability of any one of them will be increased if the remainder of the set can be assumed as given premises*' (italics in original). Tomoji Shogenji (1999) defends a coherence measure that equals the ratio of the joint probability of all the propositions in S over the product of the marginal probabilities of the propositions in S. Erik J. Olsson (2002*b*: 250) suggests as a possible measure of coherence the ratio of the joint probability of the propositions in S over the probability of the disjunction of the propositions in S. Branden Fitelson (2003) defends a measure of coherence that is based on the Kemeny and Oppenheim measure of factual support. We will return to these proposals in Section 2.6.

There is one thing that all these proposals share, viz. their probabilistic nature. We return to the information set $S = \{R_1, \ldots, R_n\}$ contained in the set of information sets $\mathbf{S}$. Let R_i be the binary propositional variable whose positive value is R_i and whose negative value is $\neg\mathrm{R}_i$ for $i = 1, \ldots, n$. The *probabilistic features* of an information set are fully expressed by the joint probability distribution over $R_1, \ldots, R_n$. We can express the second tenet that defines Bayesian Coherentism as follows:

(BC_2) A coherence ordering over $\mathbf{S}$ is fully determined by the probabilistic features of the information sets contained in $\mathbf{S}$.

So, for any two information sets $S = \{R_1, \ldots, R_m\}$ and $S' = \{R'_1, \ldots, R'_n\}$, how the coherence of S compares with the coherence of S′ is fully determined by the joint probability distribution over $R_1, \ldots, R_m$ and the joint probability distribution over $R'_1, \ldots, R'_n$. Hence, a Bayesian account of coherence should be able to offer a coherence measure *m* as a function of the probabilistic features of S and S′ so that S is no less coherent than S′ if and only if $m(S) \geq m(S')$.

For instance, suppose that we are trying to identify the culprit in a murder case. Consider the information set $S = \{R_1 = $ [The culprit is French][3], $R_2 =$ [The culprit drove away from the crime scene in a Renault]}. R_1 is the binary propositional variable whose values are R_1 and $\neg\mathrm{R}_1$. We assume that, in a population of suspects who stand an equal chance of being culprits, the French are in a minority, but most of the French drive Renaults and Renaults are rarely driven by anyone who is not French. Then we may well have the following joint probabilities: $P(R_1, R_2) = .10, P(R_1, \neg R_2) = .01, P(\neg R_1, R_2) = .01$ and $P(\neg R_1, \neg R_2) = .88$. Intuitively, this information set is highly coherent. Suppose on the other hand that we are dealing with the information set $S' = \{R'_1 =$ [The culprit is French], $R'_2 =$ [The culprit is a Presbyterian]}. We assume that the French and the Presbyterians are both minorities in our population of suspects and that French Presbyterians are very rare indeed. Then the following joint probabilities may hold: $P(R'_1, R'_2) = .01,\ P(R'_1, \neg R'_2) = .10, P(\neg R'_1, R'_2) = .10, P(\neg R'_1, \neg R'_2) = .79$. Intuitively, this information set is strongly incoherent. A measure of coherence should determine whether S ranks higher in the coherence ordering than S′. As an illustration we calculate Shogenji's measure m_s for both information sets. Remember that m_s is the ratio of the joint probability of the propositions in the information set over the product of their marginal probabilities. Thus, in our example:

$$(1.1) \qquad m_s(S) = \frac{P(R_1, R_2)}{P(R_1)P(R_2)} = \frac{.10}{(.10 + .01)(.10 + .01)} \approx 8.26,$$

$$m_s(S') = \frac{P(R'_1, R'_2)}{P(R'_1)P'(R'_2)} = \frac{.01}{(.01 + .10)(.01 + .10)} \approx .826.$$

Since $m_s(S) > m_s(S')$, the Shogenji measure squares with our intuitive ranking of coherence in this case.

We take (BC_1) and (BC_2) not only to be the core of Bayesian Coherentism, but also to have a certain independent plausibility. However,

[3] Following Quine (1960: 168), we use square brackets to refer to the proposition expressed by the enclosed sentence.

we will present an impossibility result to the effect that these theses cannot jointly be true.

1.3. MODELLING INFORMATION GATHERING

Suppose that there are n independent and partially reliable sources and that each source i informs us of a proposition R_i, for $i = 1, \ldots, n$, so that the information set is $\{R_1, \ldots, R_n\}$. Let us call R_i a *fact variable* and $REPR_i$ a *report variable*. $REPR_i$ can take on two values, viz. $REPR_i$ and $\neg REPR_i$. $REPR_i$ is the proposition that, after consultation with the proper source, there is a report to the effect that R_i is the case. $\neg REPR_i$ is the proposition that, after consultation with the proper source, there is no report to the effect that R_i is the case. We construct a probability measure P over $R_1, \ldots, R_n, REPR_1, \ldots,$ and $REPR_n$, satisfying the constraint that the sources are partially reliable and independent.[4]

For the coherence of the reports to be of any consequence, the witnesses must be partially reliable. The chance that the reports of fully reliable witnesses are false is nil, i.e. $P(REPR_i|\neg R_i) = 0$ for $i = 1, \ldots, n$. Our degree of confidence is raised to certainty in whatever fully reliable witnesses report, regardless of the degree of coherence of these reports. On the other hand, fully unreliable witnesses pay no attention whatsoever to the facts on which they are reporting. It is as if they flip a coin or cast a die to determine what they will say. Let the true positive rate be $p_i := P(REPR_i|R_i)$ and let the false positive rate be $q_i := P(REPR_i|\neg R_i)$. Then, for fully unreliable witnesses, $p_i = q_i$ for $i = 1, \ldots, n$. Clearly, the reports of fully unreliable witnesses should be of no consequence to our degree of confidence regarding the matters attested to, regardless of the coherence of the reports.[5] Hence, we stipulate that the witnesses in which we are interested here should be more informative than fully unreliable witnesses yet less informative than fully reliable witnesses,

[4] Our model of partially reliable sources matches interpretation (ii) of 'dubious information-gathering processes' in Bovens and Olsson (2000: 698). Our model of independent sources can be found in Bovens and Olsson (2000: 690 and 696–70 and 2002: 143–4) and in Earman (2000: 56–9).

[5] The reader might ask why fully unreliable witnesses are not modelled as consistent liars, i.e. $P(REPR_i|R_i) = 0$ and $P(REPR_i|\neg R_i) = 1$. The information of consistent liars is actually a very reliable guide to belief formation. We simply need to turn around the truth-value of the report to get to the truth of the matter.

i.e. $p_i > q_i > 0$ for $i = 1, \ldots, n$. To keep things simple, let us assume that all witnesses are equally reliable, i.e. $p_i = p$ and $q_i = q$ for $i = 1, \ldots, n$.[6] Hence, to model partially reliable witnesses we impose the following constraint on P:

$$p > q > 0. \quad (1.2)$$

There are two aspects to the independence of the sources to consider. First, the coherence of the reports is of little consequence when the witnesses have based their reports on information that was communicated between themselves or when they have inferred their reports from facts other than those that they are reporting on. Consider a variation on our earlier example. One witness informs us that the culprit had a French accent and the other witness informs us that the culprit drove off in a Renault. Supposing that most French have French accents and drive Renaults and few non-French have French accents or drive Renaults, the coherence of the witness reports provides a strong boost to our degree of confidence that the culprit is a Renault driver with a French accent. It would indeed be a remarkable coincidence to receive independent witness reports that fit together so well. But the coherence of these reports would be of little consequence if one witness had told the other witness that the culprit drove off in a Renault and the latter had inferred from this information that the culprit had a French accent. The coherence of the reports would also be of little consequence if both sources saw the culprit drive off in what they took to be a Renault and one of the witnesses had inferred from this that the culprit had a French accent.[7] Independent witnesses are supposed to gather information by, and only by, observing the facts they report on. They may not always provide a correct assessment of these facts, but they are not supposed to be influenced by the reports of the other witnesses, nor by the facts on which other witnesses report.

[6] We will show that Bayesian Coherentism is false even in the simple case in which the witnesses are equally reliable. So there is little point in investigating the more complex case involving unequal reliability levels for the various witnesses.

[7] Of course if the witnesses independently observed that the witness drove off in what they took to be a Renault, then the coherence of their observations would increase our confidence that the culprit is a Renault driver. But the coherence of their reports as such would not increase our confidence that the culprit is a Renault driver with a French accent.

We provide the following probabilistic interpretation of what constitutes independent witnesses. Let there be a certain chance p that we will receive a report to the effect that the culprit has a French accent given that he does indeed have a French accent and a certain chance q that we will receive a report to the effect that the culprit has a French accent given that he does not have a French accent. p and q reflect how skilful the witness is at recognizing French accents. Now suppose that we come to learn that the culprit was driving a Renault or we come to learn that another witness reported that the culprit was driving a Renault. Since the independent witness who reported on the culprit's accent strictly attended to the culprit's accent, not to the culprit's car or reports about the car, p and q remain unaffected. This stipulation translates into the following constraint on $\boldsymbol{P}$. All report variables $REPR_i$ are probabilistically independent of all the other fact variables R_j and all the other report variables $REPR_j$, given the fact variable R_i for $i = 1, \ldots, n$. In formal notation:

(1.3) $$REPR_i \perp\!\!\!\perp R_1, REPR_1, \ldots, R_{i-1}, REPR_{i-1}, R_{i+1}, REPR_{i+1}, \ldots, R_n, REPR_n | R_i, \text{ for } i = 1, \ldots, n.$$ [8]

Or equivalently, in the terminology of the theory of probabilistic causality, we say that R_i *screens off* $REPR_i$ from all other fact variables R_j and from all other report variables $REPR_j$.[9]

The degree of confidence in the information set is the conditional joint probability of the propositions in the information set, given that all the reports have come in, i.e. $\boldsymbol{P}(\mathrm{R}_1, \ldots, \mathrm{R}_n | \mathrm{REPR}_1, \ldots, \mathrm{REPR}_n)$. For simplicity, we suppress the conditionalization by introducing the posterior probability function $\boldsymbol{P}^*$:

(1.4) $$\boldsymbol{P}^*(\mathrm{R}_1, \ldots, \mathrm{R}_n) = \boldsymbol{P}(\mathrm{R}_1, \ldots, \mathrm{R}_n | \mathrm{REPR}_1, \ldots, \mathrm{REPR}_n).$$

Some additional notational conventions will permit a simple representation of this posterior probability. First, it will prove useful to define a parameter $r := 1 - q/p$ which characterizes the reliability of the witness with respect to the report in question.[10] This reliability parameter is a continuous and strictly decreasing function of the *likelihood ratio*[11] q/p—i.e. the proportion of false positives over true positives. The greater this ratio is, the less reliable the witness report is. For fully unreliable witnesses, the false positive rate q equals the true positive rate p and r takes on the value 0, whereas for fully reliable witnesses the false positive rate q equals 0 and r takes on the value 1. Since the witnesses we consider are neither fully reliable nor fully unreliable, r ranges over the open interval (0, 1).

Note that r measures the reliability of the witness *with respect to the report in question* and not the reliability of the witness *tout court*. To see this distinction, consider the case in which q equals 0. In this case, r reaches its maximal value 1, no matter what the value of p is. Certainly, a witness who provides fewer rather than more false negatives, as measured by $1 - p$, is a more reliable witness *tout court*. But when q is 0, the reliability of the witness *with respect to the report in question* is not affected by the value of $p > 0$. No matter what the value of p is, we can be fully confident that what the witness says is true, since $q = 0$—i.e. she never provides any false positives. We will use the elliptical expression of witness reliability to stand for the reliability of the witness *with respect to the report in question*, not for the reliability of the witness *tout court*.

Second, we define parameters a_i for $i = 0, \ldots, n$. Remember that a fact variable R_j can take on a positive value R_j or a negative value $\neg\mathrm{R}_j$ for $j = 1, \ldots, n$. a_i is the sum of the joint probabilities of all combinations of i negative values and $n - i$ positive values of the variables $R_1, \ldots, R_n$. For example, for an information triple containing the propositions $\mathrm{R}_1, \mathrm{R}_2$, and R_3, $a_2 = \boldsymbol{P}(\neg\mathrm{R}_1, \neg\mathrm{R}_2, \mathrm{R}_3) + \boldsymbol{P}(\neg\mathrm{R}_1, \mathrm{R}_2, \neg\mathrm{R}_3) + \boldsymbol{P}(\mathrm{R}_1, \neg\mathrm{R}_2, \neg\mathrm{R}_3)$. That is, a_2 is the sum of the joint probabilities of all combinations with two negative values and one positive value. Call $< a_0, \ldots, a_n >$ the *weight vector* of the information set $\mathbf{S} = \{\mathrm{R}_1, \ldots, \mathrm{R}_n\}$ and note that $\sum_{i=0}^{n} a_i = 1$.

We show in Appendix A.1 that, given the constraints on $\boldsymbol{P}$ in (1.2) and (1.3), the following relationship holds:

[8] This notation was introduced by Dawid (1979) and has become standard notation. See Pearl (2000) and Spirtes *et al.* (2000).

[9] See Reichenbach (1956) and Salmon (1998).

[10] We will later show that the results in Chapters 1 and 2 do not depend on this particular choice of a reliability measure.

[11] One needs to be careful when talking about the likelihood ratio in Bayesian confirmation theory. Sometimes the likelihood ratio is defined as above (e.g. in Howson and Urbach 1993: 29), sometimes as the reciprocal p/q (e.g. in Pearl 1997: 34).

$$P^*(R_1, \ldots, R_n) = \frac{a_0}{\sum_{i=0}^{n} a_i \bar{r}^i}. \tag{1.5}$$

in which $\bar{r} := 1 - r = q/p$. For a better understanding of this formula, consider the following two situations. In the first situation we choose three propositions that are the positive values of independent propositional variables. The witnesses tell us, respectively, that the culprit (R_1) was a woman, (R_2) had a Danish accent, and (R_3) drove a Ford. Suppose that our population of suspects is composed so that learning that one or two of these propositions are true (or false) does not change the probability of the other proposition(s). The diagram in Figure 1.1 depicts a possible joint probability distribution over the variables R_1, R_2, and R_3 and presents the corresponding values for a_i, for $i = 0, \ldots, 3$. In the second situation we choose three equivalent propositions. The witnesses tell us, respectively, that the culprit (R'_1) was wearing Coco Chanel shoes, (R'_2) had a French accent, and (R'_3) drove a Renault. Our population of suspects is composed so that all and only people with French accents wear Coco Chanel shoes, and all and only people who wear Coco Chanel shoes drive Renaults. The diagram in Figure 1.2 depicts a possible joint probability distribution over the variables R'_1, R'_2, and R'_3 and presents the corresponding values for a'_i, for $i = 0, \ldots, 3$. Note that $a_0 = a'_0$: The prior joint probability of the information in the former information set is equal to the prior

$a_0 = 0.064$
$a_1 = 3 \times 0.096 = 0.288$
$a_2 = 3 \times 0.144 = 0.432$
$a_3 = 0.216$

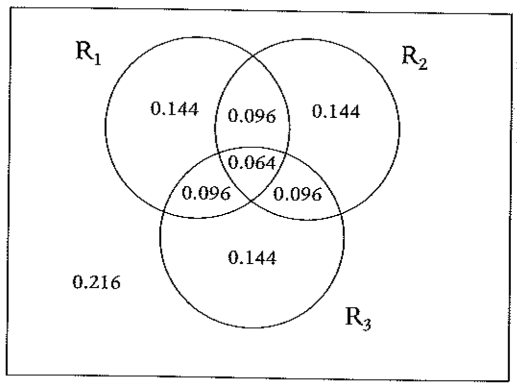

FIG. 1.1 A diagram of the joint probability distribution over the variables R_1, R_2, and R_3

$a'_0 = 0.064$
$a'_1 = 0$
$a'_2 = 0$
$a'_3 = 0.936$

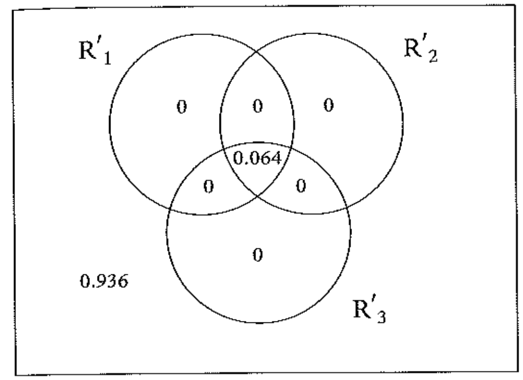

FIG. 1.2 A diagram of the joint probability distribution over the variables R'_1, R'_2, and R'_3

joint probability of the information in the latter information set. In other words, the information is equally expected in both situations. Suppose that the sources are twice as likely to issue a true positive report than a false positive report, i.e. $p = 2q$ and hence $r = \bar{r} = .50$. Then, by (1.5), our degrees of confidence after we have received the reports from the sources in these two situations are, respectively,

$$P^*(R_1, R_2, R_3) = \frac{.064}{.064 \times .50^0 + .288 \times .50^1 + .432 \times .50^2 + .216 \times .50^3} \approx .187 \text{ and} \tag{1.6}$$

$$P^*(R'_1, R'_2, R'_3) = \frac{.064}{.064 \times .50^0 + 0 \times .50^1 + 0 \times .50^2 + .936 \times .50^3} \approx .354. \tag{1.7}$$

Notice that for equally expected information and equally reliable sources, the posterior probability is greater in the second situation than in the first. And indeed, the information does fit together more tightly in the second situation. Hence, the comparison of these situations constitutes one example of the impact of relative coherence that is fully consistent with the tenets of Bayesian Coherentism. But we will now show that Bayesian Coherentism does not hold in general.

1.4. AN IMPOSSIBILITY RESULT

To disprove Bayesian Coherentism, it will suffice to construct a single counter-example to (BC_1) and (BC_2). Note that for information triples,

$a_0 + a_1 + a_2 + a_3 = 1$. Hence, from (1.5), our degree of confidence that the reports from three independent information sources are true is given by the formula:

$$(1.8) \qquad P^*(R_1, R_2, R_3) = \frac{a_0}{a_0 + a_1\bar{r} + a_2\bar{r}^2 + (1 - a_0 - a_1 - a_2)\bar{r}^3}.$$

Now pick any two information triples S and S′ with joint probability distributions that yield the respective weight vectors $<a_0, a_1, a_2, a_3> = <.05, .30, .10, .55>$ and $<a'_0, a'_1, a'_2, a'_3> = <.05, .20, .70, .05>$. The posterior joint probability that the information is true is plotted in Figure 1.3.

We do not know how to construct a coherence measure for information triples. But this does not matter. Our strategy will be to show that *any* coherence measure would leave (BC_1) and (BC_2) vulnerable to counter-examples. Hence, no reasonable proposal for a coherence measure could ever succeed.

By (BC_2), a coherence measure that induces an ordering should be a function of the probabilistic features of the information set. Since the weight vector is the only relevant information of the probability distribution in determining our degree of confidence that the information is

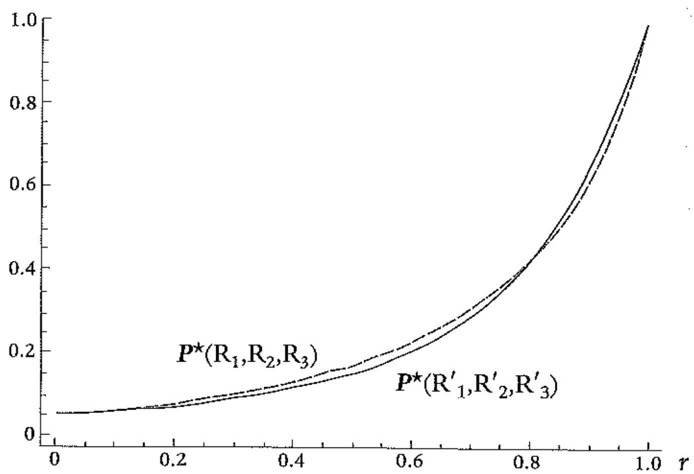

Fig. 1.3 The posterior probability for information triples with weight vectors $<a_0, a_1, a_2, a_3> = <.05, .3, .1, .55>$ and $<a'_0, a'_1, a'_2, a'_3> = <.05, .2, .7, .05>$ as a function of the reliability parameter r

true, the measure should be a function of only $<a_0, a_1, a_2, a_3>$ and $<a'_0, a'_1, a'_2, a'_3>$. However, for the present example, whatever measure we choose will violate (BC_1). To see this, first notice that the information in S and S′ is equally expected, since $a_0 = a'_0$. Suppose that we pick a measure $m_\dagger$ so that $m_\dagger(S') \geq m_\dagger(S)$. Then for any value of $r \in (.8, 1)$, (BC_1) is violated. It is not true that the more coherent the information is, the greater our degree of confidence, ceteris paribus, since $P^*(R'_1, R'_2, R'_3) < P^*(R_1, R_2, R_3)$ over this interval. Or suppose that we pick a measure $m_\ddagger$ so that $m_\ddagger(S) > m_\ddagger(S')$. Then for any value of $r \in (0, .8]$, (BC_1) is false. It is not true that the more coherent the information is, the greater our degree of confidence, ceteris paribus, since $P^*(R_1, R_2, R_3) \leq P^*(R'_1, R'_2, R'_3)$ over this interval. Thus, no measure of coherence can be constructed that determines our relative degree of confidence when all other determinants, i.e. the expectance of the information and the reliability of the witnesses, remain the same for both information sets. For the weight vectors in question, the reliability of the sources changes which information set will merit the greater degree of confidence. Similar results can be generated for information sets of size $n > 3$.[12] Hence, we can conclude that there cannot exist a measure of coherence that is probabilistic and induces a coherence ordering for information triples (BC_2) and that simultaneously makes it the case that the more coherent the information set, the more confident we are that the information is true, ceteris paribus (BC_1).

One might raise the following objections. First, we have shown that our degree of confidence is a function of the reliability r and the weight vector $<a_0, \ldots, a_n>$. It may well be the case that there is another determinant D of our degree of confidence which differs from reliability, expectance, and coherence and which is also a function of r and $< a_0, \ldots, a_n>$. (BC_1) may well be true if we keep the reliability, the expectance, *as well as* D fixed under the ceteris paribus clause. We do not have a general argument to the effect that there is no such determinant D. However, to successfully revive Bayesian Coherentism it will have to be the case that in our counter-example D has no common value in the region $r = (0, .8]$ and in the region $r = (.8, 1)$. If there are any two points in these respective regions for which D is the

[12] It is not possible to construct counter-examples of this nature for information pairs, i.e. information sets of size $n = 2$. This does not mean that Bayesian Coherentism is true for information pairs. In Chapter 2, we will show that there also does not exist a coherence ordering over the set of information pairs, violating (BC_2).

same, our counter-example will continue to apply. We cannot begin to respond to this objection without a hint of what such a determinant D might be.[13]

Second, one might object that our result is an artefact of the specific choice of the reliability measure r. However, our result holds for any measure in the class of measures that are a continuous and a strictly monotonically decreasing function of the likelihood ratio $x = q/p$. This is so because the curves of the posterior probability functions that criss-cross when plotted against r will also criss-cross when plotted against $x = 1 - r$, and hence against any continuous and strictly monotonically decreasing function of x that maps the interval (0, 1) onto the interval (0, 1). Note that the reliability measure only depends on x, and not also on, say, q. To see this, suppose that the measure were to depend on both x and q. We keep x constant and change the value of q (and accordingly the value of p). The witness reliability would thereby change, whereas, by (1.5), the posterior probability of the information would remain constant, which is unintuitive.

1.5. WEAK BAYESIAN COHERENTISM

How troubling should this negative result be? Curiously, our model at first seemed to leave some hope for a probabilistic account of coherence, but then we were able to show that Bayesian Coherentism does not hold up for *certain* pairs of information *triples*. This negative result hinges on a stipulation of the weight vectors of the pairs of information *triples* so that (i) the prior joint probabilities of the propositions in the information triples are the same and (ii) the curves of the posterior joint probability as a function of the reliability of the witnesses criss-cross—i.e. for some values of r the posterior joint probability of the propositions in one information set exceeds the posterior joint probability of the propositions in the other information set, and vice

[13] Furthermore, one might object that, if there exists such a determinant D, then coherence may well be a function of $<a_0, \ldots, a_n>$ *and* some other feature d of the probability distribution so that the coherence measure m is *not exclusively* a function of $<a_0, \ldots, a_n>$. d may be some marginal probability, as in the Shogenji measure. This could be so, as long as D is also a function of d, so that our degree of confidence is independent of d. For example let $P^*(R_1, \ldots, R_n) \sim cD$, with $m = d\, g_1(<a_0, \ldots, a_n>)$ and $D = d^{-1} g_2(<a_0, \ldots, a_n>)$. But once again, what determinant D could qualify for this role?

versa for other values of r. But note that this criss-crossing of curves does not occur for *all* pairs of information triples that satisfy condition (i). It will be instructive to compare concrete examples of pairs of information triples in which this criss-crossing occurs with examples of pairs of information triples in which this criss-crossing does not occur.

First, let us consider a case in which this criss-crossing does not occur. We return to the information triples in Section 1.3, viz. $S = \{R_1 =$ [the culprit was a woman], $R_2 =$ [the culprit had a Danish accent], $R_3 =$ [the culprit drove a Ford]$\}$ and $S' = \{R'_1 =$ [the culprit was wearing Coco Chanel shoes], $R'_2 =$ [the culprit had a French accent], $R'_3 =$ [the culprit drove a Renault]$\}$. Suppose that, given background information about the suspects, the weight vectors are respectively $<a_0, \ldots, a_3> = <.064, .288, .432, .216>$ and $<a'_0, \ldots, a'_3> = <.064, 0, 0, .936>$. Figure 1.4 plots the posterior joint probability that the information in each respective triple is true as a function of the reliability measure. Notice that these curves do not criss-cross. Hence, our degree of confidence in the information content of S is greater than our degree of confidence in the information content of S′, no matter at what level we fix the degree of partial reliability of the witnesses. For this

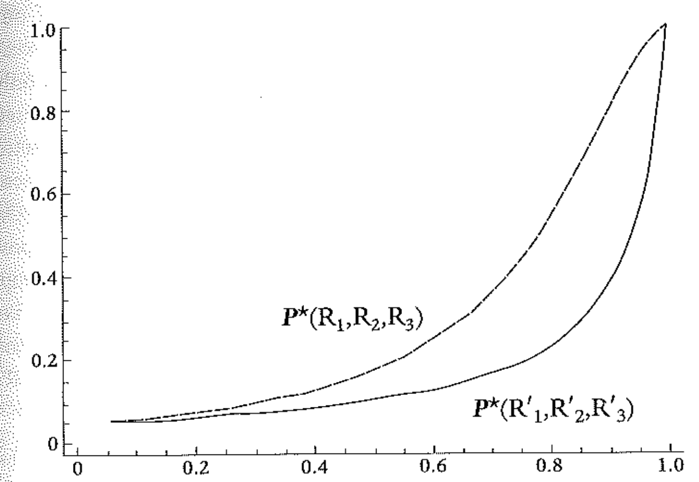

Fig. 1.4 The posterior probability for information triples with weight vectors $<a_0, a_1, a_2, a_3> = <.064, .288, .432, .216>$ and $<a'_0, a'_1, a'_2, a'_3> = <.064, 0, 0, .936>$ as a function of the reliability parameter r

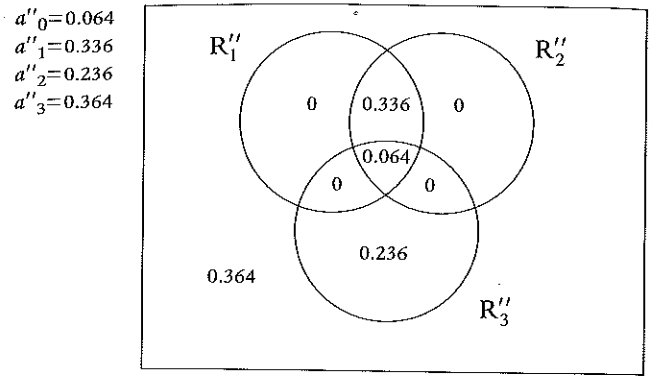

FIG. 1.5 A diagram of the joint probability distribution over the variables R_1'', R_2'', and R_3''

pair of information triples, the tenets of Bayesian Coherentism succeed.

But now compare the following two information triples. The triple S is as before, but the triple $S'' = \{R_1'' =$ [the culprit was wearing Coco Chanel shoes], $R_2'' =$ [the culprit had a French accent], $R_3'' =$ [the culprit drove a Ford]$\}$. Suppose that in our population 40 per cent of suspects wear Coco Chanel shoes, and all and only suspects who wear Coco Chanel shoes have a French accent. Also, 30 per cent of our suspects drive Fords. However, 84 per cent of the suspects who wear Coco Chanel shoes and have a French accent drive Renaults and only the remaining 16 per cent of them drive Fords. We have represented the probability distribution for S'' and calculated the weight vector $<.064, .336, .236, .364>$ in the diagram in Figure 1.5.[14] Which of these two information sets has the appearance of being more coherent? On the one hand, one might say that S is less coherent, since the propositions in S are probabilistically independent whereas S'' has two propositions, viz. R_1'' and R_2'', that are maximally positively relevant, i.e. they pick out coextensive sets of suspects. On the other hand, one might say that S is more coherent, since S'' contains a proposition R_3''

[14] Observe in the diagram that $.336 + .064 = .40$; 40% of the suspects wear Coco Chanel shoes and have a French accent; $.236 + .064 = .30$; 30% of the suspects drive Fords; $.064/.40 = .16$; 16% of the suspects with Coco Chanel shoes and French accents drive Fords. We assume that the remaining 84% of the suspects with Coco Chanel shoes and French accents drive Renaults: $.336/.40 = .84$.

that is highly negatively relevant with respect to the other propositions R_1'' and R_2''. S'' seems more coherent than a set of three independent propositions, since R_1'' and R_2'' fit together so well. Yet S'' also seems less coherent, since R_3'' fits so poorly with R_1'' and R_2''. We want to say that there just is no fact of the matter as to whether S is more or less coherent than S''. No coherence ranking can be defined over the pair $\{S, S''\}$. And indeed, if we plot the posterior joint probability for these two information sets, then we find criss-crossing lines just as in Figure 1.3.[15]

So why is it that the conjunction of (BC_1) and (BC_2) does not generally hold? We can split up (BC_2) into two components:

(BC_2^i) The *binary relation* of '... being no less coherent than ...' over *S* is fully determined by the probabilistic features of the information sets contained in *S*.

(BC_2^{ii}) The *binary relation* of '... being no less coherent than ...' is an *ordering*.

A weakened variant of Bayesian Coherentism can be salvaged if we are willing to give up (BC_2^{ii}). *Orderings* are complete, reflexive, and transitive binary relations; *quasi-orderings* are reflexive, transitive, but not necessarily complete binary relations.[16] Our suggestion is that the Bayesian Coherentist give up on the completeness requirement on the binary relation of coherence. In other words, (BC_2) needs to be replaced by

(BC_2^*) A coherence quasi-ordering over *S* is fully determined by the probabilistic features of the information sets contained in *S*.

Let us name the conjunction of (BC_1) and (BC_2^*) 'Weak Bayesian Coherentism'. According to Weak Bayesian Coherentism, there exists a coherence quasi-ordering over *S* that is fully determined by the probabilistic features of its constituent information sets. Furthermore, if S is no less coherent than S', then our degree of confidence that S is

[15] The crossing point can be obtained analytically by solving $P^*(R_1, R_2, R_3) = P^*(R_1'', R_2'', R_3'')$ for $r \in (0,1)$ using equation (1.5). This point lies at $r \approx .68$.

[16] Various terms have been used in the literature. See Sen (1970: 7–9).

true is no less than our degree of confidence that S′ is true, ceteris paribus. In our example, an ordering is defined over {S, S′} but not over {S, S″}.

How does our analysis affect the coherence theory of justification? The coherence theory is meant to be a response to Cartesian scepticism. The Cartesian sceptic claims that we are not justified in believing the story about the world that we have come by through various information-gathering processes (our senses, witnesses, etc.), since we have no reason to believe that these processes are reliable. There are many variants of the coherence theory of justification. We are interested in versions that hinge on the claim that it is the very coherence of the story of the world that gives us a reason to believe that the story is likely to be true. This is not the place to defend a full-fledged version of the coherence theory of justification, but we will argue that the substitution of (BC_2^*) for (BC_2) is not damaging to this claim.

First consider the following analogy. Suppose that we establish that the more a person reads, the more cultured she is, ceteris paribus. We conclude from this that if we meet with a very well-read person, then we have a reason to believe that she is cultured. It may not be sufficient reason, but it is a reason nonetheless. Now suppose that we also establish that sometimes no comparison can be made between the amount of reading two people do, since reading comes in many shapes and colours. We can only establish a quasi-ordering over a set of persons according to how well read they are. This does not stand in the way of our conclusion.

We have shown that, as long as our sources are independent and partially reliable, the more coherent an information set is, the more likely its content is to be true, ceteris paribus. We conclude from this that, if the story of the world is a very coherent information set, then we have a reason to believe that its content is likely to be true. Again, it may not be sufficient reason, but it is a reason nonetheless. And similarly, the fact that we can only establish a coherence quasi-ordering over information sets does not stand in the way of this conclusion.

What is misguided in the coherence theory of justification is the persistent attempt to construct a measure that imposes an ordering on sets of information sets. The coherence theory is thought to be lacking unless we have a clear measure of coherence that permits us to order information sets. What we have shown is that the insistence on such a measure is wrong-headed, since there simply is no such measure that

also respects (BC_1). A coherence theory that draws on a probabilistic measure of coherence must make do with a quasi-ordering.

Opponents of the coherence theory may try to get some mileage out of our result. Indeed, a radical response to what has been demonstrated would be to discard the Coherentism part of Bayesian Coherentism—i.e. (BC_1). But our formal model actually discourages this move. Our model shows that if S is indeed more coherent than S′, then our degree of confidence in the content of S should be greater than in the content of S′, assuming that the sources are equally reliable and the information is equally expected. The opponent of the Coherentism part of Bayesian Coherentism will need to show that this formal model is not fit to deal with Cartesian scepticism.

A more moderate response to our analysis would be to tinker with (BC_2), i.e. with the Bayesian part of Bayesian Coherentism. We have proposed altering (BC_2^{ii}), but others have tried to tinker with (BC_2^{i}). The idea underlying such proposals is that probabilistic accounts of coherence cannot do justice to the richness of this concept, and that negative results are to be expected when one works within such an impoverished structure. For instance, it has been argued that the coherence of an information set should take into account explanatory relations between the propositions in the set, and that these relations cannot be adequately represented by probabilistic information (e.g. BonJour 1985: 99–101). If one takes into account the full richness of the notion of coherence, then it is possible to construct an ordering—and hence to respect (BC_2^{ii})—of information sets, or so the argument goes. At this point in time we may not have a principled way of doing so, but proponents argue that *this* is the challenge that the coherence theory of justification must take up. We take no dogmatic stand on this issue, but remain suspicious of any claim to the effect that there are aspects of uncertain reasoning that resist a strictly probabilistic analysis.

2

Coherence

2.1. UNEQUAL PRIORS

In the previous chapter we showed that there cannot be a measure that induces a coherence ordering—i.e. a binary relation which is complete, reflexive, and transitive—over the set of possible information sets. This does not exclude the construction of a measure that induces a coherence quasi-ordering—i.e. a binary relation which is reflexive and transitive. So far we have only considered a special case—we have laid out a procedure to order pairs of equal-sized information sets that share the same prior probability that their respective constitutive propositions are all true. In effect, we have partitioned the set of all information sets into subsets **S** of information sets that have the same cardinality and the same prior joint probability a_0. Within each of these subsets **S** we have constructed a procedure to impose a quasi-ordering over **S**. Let '$\succeq$' be the binary relation of *being no less coherent than*. Then for pairs of information sets $S = \{R_1, \ldots, R_n\}$ and $S' = \{R'_1, \ldots, R_n'\}$, our procedure can be stated as follows:

(2.1) For all S, $S' \in \mathbf{S}$, if S and S' have the same cardinality and $\mathbf{P}(R_1, \ldots, R_n) = a_0 = a'_0 = \mathbf{P}(R'_1, \ldots, R'_n)$, then $S \succeq S'$ iff $\mathbf{P}^*(R_1, \ldots, R_n) \geq \mathbf{P}^*(R'_1, \ldots, R'_n)$ for all values of the reliability parameter $r \in (0, 1)$.

In other words, S is no less coherent than S' if and only if the curve representing the function for their posterior joint probability for S is strictly above the curve for S' over the interval $r \in (0, 1)$. We have assumed that the witnesses are equally reliable and will discuss this assumption in Section 2.4.

We should be able to do better than this. Our intuitive notion of one information set being no less coherent than another information set is not restricted to information sets whose content is equally probable nor to information sets of the same cardinality. Let us look at a few examples.

First, suppose that a murder has been committed in Tokyo. We are trying to locate the corpse and, given our background knowledge, every square inch of Tokyo is just as likely a spot as every other square inch. Suppose two witnesses independently point to a particular house. This is certainly coherent information. Alternatively, suppose that one witness points to some broad area on the map and the other witness points to an area that is no less broad. The overlap between both areas is a large district of Tokyo. There is little doubt that the information in the first case is more coherent than the information in the second case. And yet the prior probability that the information of the witnesses in the first case is true is much lower than the prior probability that the information of the witnesses in the second case is true, for the house is a much smaller region than the district.

Second, BonJour poses the following example of information sets that can clearly be ordered with respect to their relative coherence. Consider the following two information sets: S = {[All ravens are black], [This bird is a raven], [This bird is black]} and S' = {[This chair is brown], [Electrons are negatively charged], [Today is Thursday]} (1985: 96). There is no doubt that set S is more coherent than set S'. And yet there is no reason to assume that the prior probability that the information in S is true equals the prior probability that the information in S' is true.

Third, we also make judgements of relative coherence when the information sets are of unequal size. For instance, consider the paradigm case of non-monotonic reasoning. Certainly the information *pair* S = {[My pet Tweety is a bird], [My pet Tweety cannot fly]} is less coherent than the information *triple* S' = {[My pet Tweety is a bird], [My pet Tweety cannot fly], [My pet Tweety is a penguin]}. The inclusion of the information that Tweety is a penguin is what brings coherence to the story. What we want is a measure that induces a coherence quasi-ordering over information sets in general, not just information sets of the same size and with equal prior joint probabilities.

Various attempts have been made to provide a probabilistic account of the notion of coherence. In the previous chapter we showed that the search for a measure that imposes a coherence ordering on the set of information sets is in vain. However, a coherence quasi-ordering should suffice for the purposes of the coherence theory of justification. Thus, in this chapter, we will take on the project of showing how to construct a general measure that imposes a coherence quasi-ordering on the set of information sets.

The notion of coherence also plays a role in philosophy of science. Kuhn (1977: 321–2, quoted in Salmon (1990: 176)) mentions *consistency* as one of the (admittedly imprecise) criteria for scientific theory choice (along with accuracy, scope, simplicity, and fruitfulness). Salmon (1990: 198) distinguishes between the internal consistency of a theory and the consistency of a theory with other accepted theories. In discussing the latter type of consistency, he claims that there are two aspects to this notion, viz. the '*deductive* relations of entailment and compatibility' and the '*inductive* relations of fittingness and incongruity'. We propose to think of the internal consistency of a theory in the same way as Salmon thinks of the consistency of a theory with accepted theories. Hence, the *internal consistency* of a theory matches the epistemologist's notion of the *coherence* of an information set: How well do the various components of the theory fit together, how congruous are these components? Salmon also writes that this criterion of consistency 'seem[s] ... to pertain to assessments of the prior probabilities of the theories' and 'cr[ies] out for a Bayesian interpretation' (1990: 198). Following this line of thought, we will show how one can construct a coherence quasi-ordering over a set of scientific theories and how our relative degree of confidence that one or another scientific theory is true is functionally dependent on this quasi-ordering. That the relation is a quasi-ordering rather than an ordering respects Kuhn's contention that consistency is an imprecise criterion of theory choice. Indeed, in some cases, it is indeterminate which of two theories is more coherent.

2.2. CONSTRUCTING A MEASURE

We will construct a formal measure that permits us to read off a coherence quasi-ordering from the joint probability distributions over the propositional variables whose positive values are constitutive of the information sets. 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. This is a mistake. To see this, consider the following analogy. We not only use the notion of coherence when we talk about information sets, but also, for example, when we talk about groups of individuals. Group coherence tends to be a good thing. It makes ant colonies more fit for survival, it makes law firms more efficient, it makes for happier families, etc. It makes little sense to ask what makes for a more coherent group independently of the particular role that coherence is supposed to play in the context in question. We must first fix the context in which coherence purports to play a particular role. For instance, let the context be ant colonies and let the role be that of promoting reproductive fitness. We give more precise content to the notion of coherence in this context by letting coherence be the property of ant colonies that plays the role of boosting fitness and at the same time matches our pre-theoretic notion of the coherence of social units. A precise fill-in for the notion of coherence will differ as we consider fitness boosts for ant heaps, efficiency boosts for law firms, or happiness boosts for families.

Similarly, it makes little sense to ask precisely what makes for a more coherent information set independently of the particular role that coherence is supposed to play. The coherence theory of justification and the Kuhnian appeal to coherence as a criterion of theory choice ride on a particular common-sense intuition. When we gather information from independent and partially and equally reliable sources, the more coherent the story is, the more confident we are that the story is true, ceteris paribus. Within the context of information gathering from such sources, coherence is a property of information sets that plays a confidence-boosting role.

In the previous chapter we derived a parsimonious expression for the posterior probability that the information is true which we receive from independent witnesses who are partially and equally reliable:

$$P^*(\mathrm{R}_1, \ldots, \mathrm{R}_n) = \frac{a_0}{\sum_{i=0}^{n} a_i \bar{r}^i}. \tag{2.2}$$

Remember that $\bar{r} := 1 - r$, with r being the reliability parameter equal to $1 - q/p$. The true positive rate $p := P(\mathrm{REPR_i}|\mathrm{R_i})$ is greater than the false positive rate $q := P(\mathrm{REPR_i}|\neg\mathrm{R_i})$ which is greater than 0 for

$i = 1, \ldots, n$. $<a_0, \ldots, a_n>$ is the weight vector of the information set S = {$R_1, \ldots, R_n$}. Each a_i is the sum of the joint probabilities of all combinations of i negative values $\neg R_j$ and $n - i$ positive values R_j of the propositional variables $R_1, \ldots, R_n$.

A maximally coherent information set has the weight vector $<a_0, 0, \ldots, 0, \bar{a}_0>$ with $\bar{a}_0 := 1 - a_0$. Let us assume that we are neither certain that the content of the information set is true nor certain that it is false. All items of information $R_1, \ldots, R_n$ are equivalent, since $a_0 = P(R_1, \ldots, R_n)$ and $\bar{a}_0 = a_n = P(\neg R_1, \ldots, \neg R_n)$ and the joint probabilities of all other combinations of propositions are set at 0. If one of the remaining $a_1, \ldots$, or a_{n-1} exceeds 0, then the items of information are no longer equivalent and the information set loses its maximal coherence. It is some feature of $<a_0, \ldots, a_n>$ that determines the coherence of the information set. For maximal coherence, it needs to be the case that $a_i = 0$ for $i = 1, \ldots, n - 1$. But it is not clear at all what feature we are looking for when assessing and comparing cases of non-maximal coherence.

To determine this feature, here is how we will proceed. Suppose that we have a range of suspects for some crime. We question the witnesses, who provide us information about what car the culprit was driving, the culprit's accent, etc. All this information picks out a certain subset of the original suspects that satisfy all these features. Let's suppose that only Jean and Pierre satisfy these features. The information that led us to pick out Jean or Pierre may have been maximally coherent. For instance, it may be the case that each witness provided a report that it was either Jean or Pierre who was the culprit. Or it may be the case that one witness claimed that the culprit is from Marseille and the other witness claimed that the culprit is a sailor and that all and only inhabitants from Marseille are sailors in our population of suspects. But the information may also have been less coherent. One witness might have said that the suspect had a French accent and the other witness that the suspect was a Presbyterian. The population of suspects contains a large subset of suspects with French accents and a large subset of suspects who are Presbyterians, but only Jean and Pierre are Presbyterians with French accents. We learned in the last chapter that for any particular value of the reliability parameter r, our confidence boost that either Jean or Pierre is the suspect is much greater when the information comes to us in the form of maximally coherent information rather than in the form of less than maximally coherent information. Our strategy will be to assess the coherence of an information set by measuring the proportion of the confidence boost that we actually receive, relative to the confidence boost that we would have received *had we received this very same information in the form of maximally coherent information.*

To put this formally, let us turn to our example of independent tests that identify sections on the human genome that may contain the locus of a genetic disease. The tests pick out different areas, and the overlap between the areas is a region σ. The information is more coherent when the reports are all clustered around the region σ than when they are scattered all over the human genome but have this relatively small area of overlap on the region σ. The information is maximally coherent when every single test points to the region σ. We assign a certain prior probability that the locus of the disease is in the region σ. With more coherent reports, our confidence boost will be greater than with less coherent reports. Let us measure this confidence boost by the ratio of the posterior probability—i.e. the probability after we have received the reports—over the prior probability that the locus of the disease is in region σ:

$$b(\{R_1, \ldots, R_n\}) = \frac{P^*(R_1, \ldots, R_n)}{P(R_1, \ldots, R_n)}. \tag{2.3}$$

To determine this confidence boost it is sufficient to know the weight vector $<a_0, \ldots, a_n>$ and the reliability parameter r, since $P(R_1, \ldots, R_n)$ equals a_0 and since $P^*(R_1, \ldots, R_n)$ is a function of the weight vector and the reliability parameter.

If we had received the information that the locus of the disease is in region σ in the form of maximally coherent information, then our information set would have contained n reports to the effect that the locus of the disease was in region σ, i.e. $\{R_1^\sigma, \ldots, R_n^\sigma\}$. We can impose a probability measure P^{max} over the propositional variables $R_1^\sigma, \ldots, R_n^\sigma$ with the corresponding weight vector $<a_0, 0, \ldots, 0, a_n>$. We insert this weight vector into (2.2) and calculate what our degree of confidence would have been that the locus of the disease is in region σ, had we received the information as maximally coherent information:

$$P^{max*}\left(R_1^\sigma, \ldots, R_n^\sigma\right) = \frac{a_0}{a_0 + \bar{a}_0 \bar{r}^n}. \tag{2.4}$$

Hence, our confidence boost would have been

$$b^{max}(\{R_1, \ldots, R_n\}) = \frac{P^{max*}(R_1^{\sigma}, \ldots, R_n^{\sigma})}{P^{max}(R_1^{\sigma}, \ldots, R_n^{\sigma})}. \tag{2.5}$$

Since the prior probability $P^{max}(R_1^{\sigma}, \ldots, R_n^{\sigma}) = P(R_1, \ldots, R_n) = a_0$, the proportion of the confidence boost that we actually receive, relative to the confidence boost that we would have received, had we received this very same information in the form of maximally coherent information, equals

$$\begin{aligned} c_r(\{R_1, \ldots, R_n\}) &= \frac{b(\{R_1, \ldots, R_n\})}{b^{max}(\{R_1, \ldots, R_n\})} \\ &= \frac{P^*(R_1, \ldots, R_n)/P(R_1, \ldots, R_n)}{P^{max*}(R_1^{\sigma}, \ldots, R_n^{\sigma})/P^{max}(R_1^{\sigma}, \ldots, R_n^{\sigma})} \\ &= \frac{P^*(R_1, \ldots, R_n)}{P^{max*}(R_1^{\sigma}, \ldots, R_n^{\sigma})} \\ &= \frac{a_0 + \bar{a}_0\bar{r}^n}{\sum_{i=0}^{n} a_i\bar{r}^i}. \end{aligned} \tag{2.6}$$

This measure is functionally dependent on the reliability parameter r. Clearly, our pre-theoretic notion of the coherence of an information set does not encompass the reliability of the witnesses that provide us with its content. So how can we use this measure to assess the relative coherence of two information sets?

Let us look at what we did in the special case in which information sets S and S′ have the same cardinality and $P(R_1, \ldots, R_n) = a_0 = a'_0 = P(R'_1, \ldots, R'_n)$. We salvaged the core of Bayesian Coherentism by imposing an ordering on a pair of information sets if and only if the curves representing the posterior probabilities that the contents of the information sets are true as a function of r do not criss-cross. Formally, $S \succeq S'$ if and only if $P^*(R_1, \ldots, R_n) \geq P^*(R'_1, \ldots, R'_n)$ for all values of the reliability parameter $r \in (0, 1)$. This permitted us to respect the first tenet of Bayesian Coherentism—viz. the more coherent an information set is, the greater our degree of confidence that its content is true, ceteris paribus—while remaining faithful to a weakened

version of the second tenet—viz. that the quasi-ordering of *being no less coherent than* is determined by the probabilistic features of the information set.

In the general case, we would like to be able to assess and compare the coherence of information sets that may not have the same cardinality and may not share the same joint prior probability that their respective contents are true. Our strategy is to assess the coherence of an information set by measuring the proportion of the confidence boost that we actually receive, relative to the confidence boost that we would have received, had we received this very same information in the form of maximally coherent information. Also, in the general case we would like to be able to make the claim that the more coherent an information set is, the greater this proportional confidence boost, ceteris paribus, in which the ceteris paribus clause requires that the reliability parameter r be held constant. Now we run into precisely the same problem that we ran into before: Some pairs of information sets $\{S, S'\}$ are such that $c_r(S) > c_r(S')$ for some values of r, whereas $c_r(S') > c_r(S)$ for other values of r. To safeguard our current claim, we follow the same strategy. We impose an ordering on a pair of information sets if and only if the curves that represent the proportional confidence boosts as a function of r do not criss-cross. In formal terms,

(2.7) For all $S, S' \in \mathbf{S}$, $S \succeq S'$ iff $c_r(S) \geq c_r(S')$ for all values of the reliability parameter $r \in (0, 1)$.

This procedure induces a quasi-ordering on the set of information sets in general, whatever their cardinalities and whatever the prior joint probabilities that their contents are true. We will see that this distinction squares with our willingness to make intuitive judgements about the relative coherence of information sets.

The reader may wonder whether our general-case procedure entails our special-case procedure. The answer is straightforward. In the special case, we assume that the cardinalities of the information sets are equal and that the prior probabilities that the contents of the information sets are true are equal—i.e. $a_0 = a'_0$. From (2.2) and (2.6), it follows that we can write the posterior joint probability that the content of the information set is true as follows:

$$(2.8)\qquad P^*(R_1, \ldots, R_n) = \frac{a_0}{a_0 + \bar{a}_0 \bar{r}^n} c_r(\{R_1, \ldots, R_n\}).$$

It is clear from (2.8) that

(2.9) For all S, $S' \in \mathbf{S}$, if S has cardinality m and S' has cardinality n with $m = n$ and $a_0 = a'_0$, then $P^*(R_1, \ldots, R_m) \geq P^*(R'_1, \ldots, R'_n)$ if and only if $c_r(S) \geq c_r(S')$ for all values of the reliability parameter $r \in (0, 1)$.

Our procedure in the general case, as expressed in (2.7), in conjunction with (2.9) entails our procedure in the special case, as expressed in (2.1).

Rather than assessing directly whether the curves criss-cross for the functions that measure the proportional confidence boost, we construct a *difference function*. Consider two information sets $S = \{R_1, \ldots, R_m\}$ and $S' = \{R'_1, \ldots, R'_n\}$. We calculate the weight vectors $< a_0, \ldots, a_m >$ and $< a'_0, \ldots, a'_n >$. The difference function is defined as follows:

$$(2.10)\qquad f_r(S, S') = c_r(S) - c_r(S').$$

$f_r(S, S')$ has the same sign for all values of $r \in (0, 1)$ if and only if the measure $c_r(S)$ is always greater than or is always smaller than the measure $c_r(S')$ for all values of $r \in (0, 1)$. Hence, we can restate the general procedure in (2.7) that induces a quasi-ordering over an unrestricted set of information sets in a more parsimonious fashion:

(2.11) For two information sets $S, S' \in \mathbf{S}$, $S \succeq S'$ iff $f_r(S, S') \geq 0$ for all values of $r \in (0, 1)$.

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 weight vectors $< a_0, \ldots, a_n >$ and $< a'_0, \ldots, a'_n >$. One need only evaluate the conditions under which the sign of the difference function is invariable for all values of $r \in (0, 1)$. In Appendix B.1, we have shown that

(2.12) $a'_i/a_i \geq \max(1, a'_0/a_0),\ \forall i = 1, \ldots, n-1$
is a necessary and sufficient condition for $S \succeq S'$ for $n = 2$ and is a sufficient condition for $S \succeq S'$ for $n > 2$.

This is the more parsimonious statement of the condition. However, it is easier to interpret this condition when stated as a disjunction:

(2.13) (i) $a'_0 \leq a_0$ & $a'_i \geq a_i,\ \forall i = 1, \ldots, n-1$, or,
(ii) $a'_0 \geq a_0$ & $a'_i/a_i \geq a'_0/a_0,\ \forall i = 1, \ldots, n-1$,
is a necessary and sufficient condition for $S \succeq S'$ for $n = 2$ and is a sufficient condition for $S \succeq S'$ for $n > 2$.

It is easy to see that (2.12) and (2.13) are equivalent.[1]

Let us now interpret (2.13). For $n = 2$, let $S = \{R_1, R_2\}$ and consider the diagram for the joint probability distribution in Figure 2.1. There are precisely two ways to decrease[2] the coherence in moving from information sets S to S': First, by shrinking the overlapping area between R_1 and R_2 ($a'_0 \leq a_0$) and expanding the non-overlapping area ($a'_1 \geq a_1$); and second, by expanding the overlapping area ($a'_0 \geq a_0$) while expanding the non-overlapping area to a greater degree ($a'_1/a_1 \geq a'_0/a_0$). The example of the corpse in Tokyo in the next section is meant to show that these conditions are intuitively plausible.

For $n > 2$, consider the diagram for the joint probability distribution in Figure 2.2 and let $S = \{R_1, R_2, R_3\}$. There are two ways to decrease the coherence in moving from S to S': First, by shrinking the area in which there is complete overlap between $R_1, \ldots, R_n$ ($a'_0 \leq a_0$)

[1] Assume (2.12). Either $max(1, a'_0/a_0) = 1$ or $max(1, a'_0/a_0) = a'_0/a_0$. In the former case, it follows from the inequality in (2.12) that $a'_0 \leq a_0$ and $a'_i \geq a_i,\ \forall i = 1, \ldots, n-1$. In the latter case, it follows from the inequality in (2.12) that $a'_0 \geq a_0$ and $a'_i/a_i \geq a'_0/a_0,\ \forall i = 1, \ldots, n-1$. Hence, (2.13) follows. Assume (2.13). Suppose (i) holds. From the first conjoint in (i), $max(1, a'_0/a_0) = 1$ and hence from the second conjoint in (i), $a'_i/a_i \geq max(1, a'_0/a_0), \forall i = 1, \ldots, n-1$. Suppose (ii) holds. From the first conjoint in (ii), $max(1, a'_0/a_0) = a'_0/a_0$ and hence from the second conjoint in (ii), $a'_i/a_i \geq max(1, a'_0/a_0), \forall i = 1, \ldots, n-1$. Hence, (2.12) follows.

[2] We introduce the convention that 'decreasing' stands for *decreasing or not changing*, 'shrinking' for *shrinking or not changing*, and 'expanding' for *expanding or not changing*. This convention permits us to state the conditions in (2.13) more clearly and is analogous to the microeconomic convention to let 'preferring' stand for *weak* preference, i.e. for *preferring to or being indifferent between* in ordinary language.

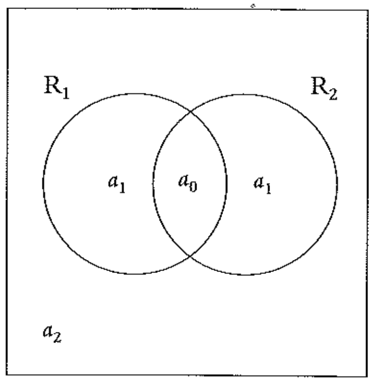

FIG. 2.1 A diagram for the probability distribution for information pairs

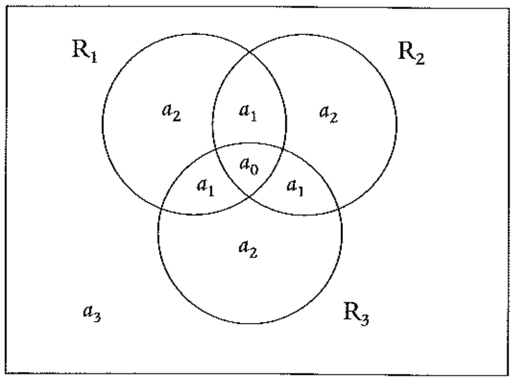

FIG. 2.2 A diagram for the probability distribution for information triples

and expanding all the areas in which there is no complete overlap ($a'_i \geq a_i, \forall i = 1, \ldots, n-1$); and second, by expanding the area in which there is complete overlap ($a'_0 \geq a_0$) and expanding all the non-overlapping areas to a greater degree ($a'_i/a_i \geq a'_0/a_0, \forall i = 1, \ldots, n-1$). This is a sufficient but not a necessary condition for $n > 2$. Hence, if equal-sized information sets do not satisfy condition (2.13), we still need to apply our general method in (2.11), i.e. we need to examine the sign of $f_r(S, S')$ for all values of $r \in (0, 1)$. The example of BonJour's challenge in the next section shows that it may be possible

to order two information sets using the general method in (2.11) without satisfying the sufficient condition in (2.13).

If we wish to determine the relative coherence of two information sets S and S′ of unequal size, we have no shortcut. In that case, we need to apply our general method in (2.11), i.e. we need to examine the sign of $f_r(S, S')$ for all values of $r \in (0, 1)$. The example of Tweety in the next section will provide an illustration of the procedure used to judge the relative coherence of information sets of unequal size.

2.3. A CORPSE IN TOKYO, BONJOUR'S RAVENS AND TWEETY

Does our analysis yield the correct results for some intuitively clear cases? We consider a comparison (i) of two information *pairs*, (ii) of two information *triples*, and (iii) of two information sets of unequal size.

(i) Information Pairs. Suppose that we are trying to locate a corpse from a murder somewhere in Tokyo. We draw a grid of 100 squares over the map of the city and consider it equally probable that the corpse lies somewhere within each square. We interview two partially and equally reliable witnesses. Suppose witness 1 reports that the corpse is somewhere in squares 50 to 60 and witness 2 reports that the corpse is somewhere in squares 51 to 61. Call this situation α and include this information in the information set S^α. For this information set, $a_0^\alpha = .10$ and $a_1^\alpha = .02$.

Let us now consider a different situation in which the reports from the two sources overlap far less. In this alternate situation—call it β—witness 1 reports squares 20 to 55 and witness 2 reports squares 55 to 90. This information is contained in S^β. The overlapping area shrinks to $a_0^\beta = .01$ and the non-overlapping area expands to $a_1^\beta = .70$. On condition (2.13)(i), S^β is less coherent than S^α, since $a_0^\beta = .01 \leq a_0^\alpha = .10$ and $a_1^\beta = .70 \geq a_1^\alpha = .02$.

In a third situation γ, witness 1 reports squares 20 to 61 and witness 2 reports squares 50 to 91. S^γ contains this information. The overlapping area expands to $a_0^\gamma = .12$ and the non-overlapping area expands to $a_1^\gamma = .60$. On condition (2.13)(ii), S^γ is less coherent than S^α, since $a_0^\gamma = .12 \geq a_0^\alpha = .10$ and $a_1^\gamma/a_1^\alpha = 30 \geq 1.2 = a_0^\gamma/a_0^\alpha$.

Now let us consider a pair of situations in which no ordering of the information sets is possible. We are considering information pairs, i.e. $n = 2$, and so condition (2.12) and (2.13) provide equivalent necessary and sufficient conditions to order two information pairs, *if there exists an ordering*. In situation δ, witness 1 reports squares 41 to 60 and witness 2 reports squares 51 to 70. So $a_0^\delta = .10$ and $a_1^\delta = .20$. In situation ε, witness 1 reports squares 39 to 61 and witness 2 reports squares 50 to 72. So $a_0^\varepsilon = .12$ and $a_1^\varepsilon = .22$. Is the information set in situation δ more or less coherent than in situation ε? It is more convenient here to invoke condition (2.12). Notice that $a_1^\varepsilon / a_1^\delta = 1.10$ is not greater than or equal to $1.20 = max\left(1, a_0^\varepsilon / a_0^\delta\right)$, nor is $a_1^\delta / a_1^\varepsilon \approx .91$ greater than or equal to $1 = max\,(1, a_0^\delta / a_0^\varepsilon)$. Hence neither $S^\delta \succeq S^\varepsilon$ nor $S^\varepsilon \succeq S^\delta$ hold true.

These quasi-orderings over the information sets in situations α and β, in situations α and γ, and in situations δ and ε seems to square quite well with our intuitive judgements. Without having done any empirical research, we conjecture that most experimental subjects would indeed rank the information set in situation α to be more coherent than the information sets in either situations β or γ. Furthermore, we also conjecture that if one were to impose sufficient pressure on the subjects to judge which of the information sets in situations δ and ε is more coherent, we would be left with a split vote.

We have reached these results by applying the special conditions in (2.12) and (2.13) for comparing information sets. The same results can be obtained by using the general method in (2.11). Write down the difference functions as follows for each comparison (i.e. let $i = \alpha$ and $j = \beta$, let $i = \alpha$ and $j = \gamma$, and let $i = \delta$ and $j = \varepsilon$ in turn):

$$(2.14)\quad f_r(S^i, S^j) = c_r(S^i) - c_r(S^j) = \frac{a_0^i + \bar{a}_0^i \bar{r}^2}{a_0^i + a_1^i \bar{r} + a_2^i \bar{r}^2} - \frac{a_0^j + \bar{a}_0^j \bar{r}^2}{a_0^j + a_1^j \bar{r} + a_2^j \bar{r}^2}.$$

As we can see in Figure 2.3, the functions $f_r(S^\alpha, S^\beta)$ and $f_r(S^\alpha, S^\gamma)$ are positive for all values of $r \in (0, 1)$—so S^α is more coherent than S^β and S^γ. But $f_r(S^\delta, S^\varepsilon)$ is positive for some values and negative for other values of $r \in (0, 1)$—so there is no coherence ordering over S^δ and S^ε.

(ii) Information Triples. We return to BonJour's challenge. There is a more coherent set, $S = \{R_1 = $ [All ravens are black], $R_2 =$ [This bird is a raven], $R_3 =$ [This bird is black]}, and a less coherent set, $S' = \{R'_1 =$ [This chair is brown], $R'_2 =$ [Electrons are negatively charged],

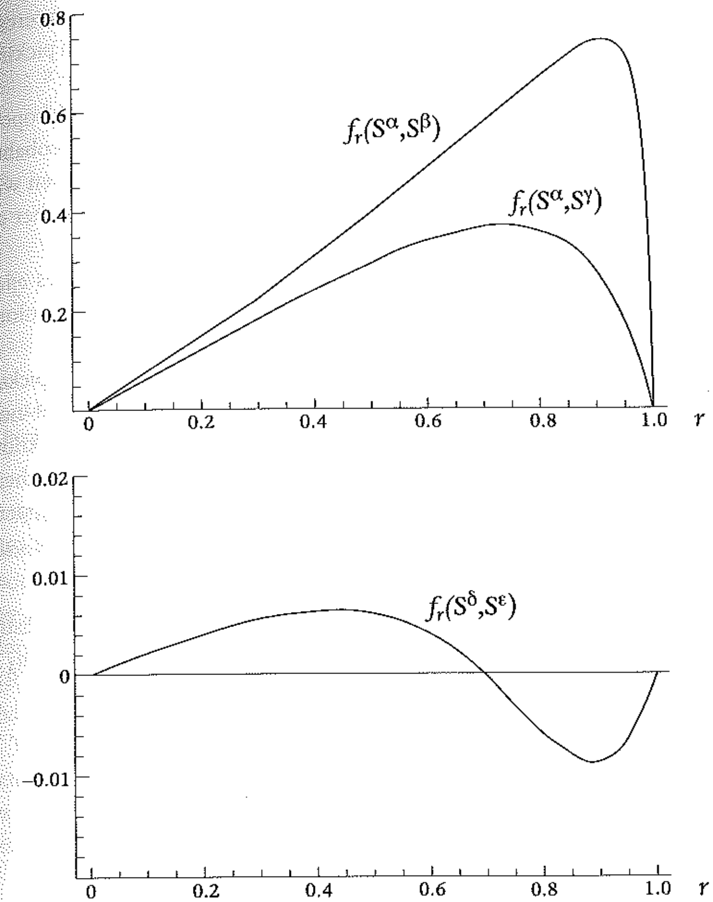

FIG. 2.3 The difference functions for a corpse in Tokyo

$R'_3 =$ [Today is Thursday]}. The challenge is to give an account of the fact that S is more coherent than S′. Let us apply our analysis to this challenge.

What is essential in S is that $R_1 \& R_2 \vdash R_3$, so that $P(R_3 | R_1, R_2) = 1$. But to construct a joint probability distribution, we need to make some additional assumptions. Let us make assumptions that could plausibly describe the degrees of confidence of an amateur ornithologist who is sampling a population of birds:

(i) There are four species of birds in the population of interest, ravens being one of them. There is an equal chance of picking a bird from each species: $P(R_2) = 1/4$.

(ii) The random variables R_1 and R_2, whose values are the propositions R_1 and $\neg R_1$, and R_2 and $\neg R_2$, respectively, are probabilistically independent: Learning no more than that a raven was (or was not) picked teaches us nothing at all about whether all ravens are black.

(iii) We have prior knowledge that birds of the same species *often* have the same colour and black may be an appropriate colour for a raven. Let us set $P(R_1) = 1/4$.

(iv) There is a one in four chance that a black bird has been picked amongst the non-ravens, whether all ravens are black or not, i.e. $P(R_3|\neg R_1, \neg R_2) = P(R_3|R_1, \neg R_2) = 1/4$. Since we know that birds of a single species often share the same colour, there is only a chance of $1/10$ that the bird that was picked happens to be black, given that it is a raven and that it is not the case that all ravens are black, i.e. $P(R_3|\neg R_1, R_2) = 1/10$.

These assumptions permit us to construct the joint probability distribution for R_1, R_2, R_3 and to specify the weight vector $<a_0, \ldots, a_3>$ (see Figure 2.4).[3]

$a_0 = 1/16$
$a_1 = 21/320$
$a_2 = 9/20$
$a_3 = 27/64$

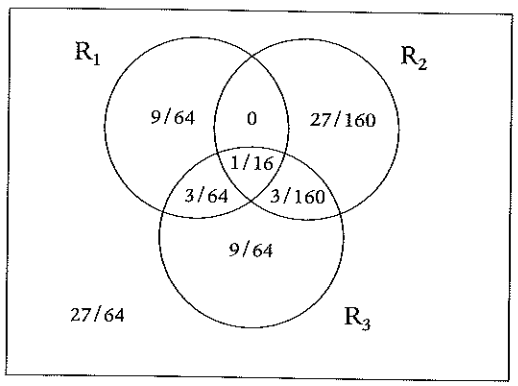

FIG. 2.4 A diagram for the probability distribution for the set of dependent propositions in BonJour's ravens

[3] Since R_1 and R_2 are probabilistically independent, $P(R_1, R_2, R_3) = P(R_1)P(R_2)P(R_3|R_1, R_2)$ for all values of R_1, R_2, and R_3. The numerical values in Figure 2.4 can be directly calculated.

$a'_0 = 1/64$
$a'_1 = 9/64$
$a'_2 = 27/64$
$a'_3 = 27/64$

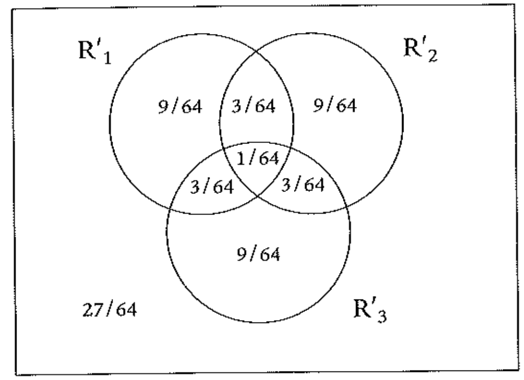

FIG. 2.5 A diagram for the probability distribution for the set of independent propositions in BonJour's ravens

What is essential in information set S′ is that the propositional variables are probabilistically independent—e.g. learning something about electrons presumably does not teach us anything about what day it is today or about the colour of a chair. Let us suppose that the marginal probabilities of each proposition are $P(R'_1) = P(R'_2) = P(R'_3) = 1/4$. We construct the joint probability distribution for R'_1, R'_2, and R'_3 and specify the weight vector $<a'_0, \ldots, a'_3>$ in Figure 2.5.[4]

The information triples do not pass the *sufficient* condition for the determination of the direction of the coherence ordering in (2.12).[5] So we need to appeal to our general method and construct the difference function:

$$(2.15)\quad f_{ravens} = f_r(S, S') = \frac{a_0 + \bar{a}_0\bar{r}^3}{a_0 + a_1\bar{r} + a_2\bar{r}^2 + a_3\bar{r}^3} - \frac{a'_0 + \bar{a}'_0\bar{r}^3}{a'_0 + a'_1\bar{r} + a'_2\bar{r}^2 + a'_3\bar{r}^3}.$$

We have plotted f_{ravens} in Figure 2.7. This function is positive for all values of $r \in (0, 1)$. Hence we may conclude that S is more coherent than S′, which is precisely the intuition of which BonJour wanted an account.[6]

[4] Since R'_1, R'_2, and R'_3 are probabilistically independent, $P(R'_1, R'_2, R'_3) = P(R'_1)P(R'_2)P(R'_3)$ for all values of R'_1, R'_2, and R'_3. The numerical values in Figure 2.5 can be directly calculated.

[5] Clearly the condition fails for $S' \succeq S$, but it also fails for $S \succeq S'$, since $a'_2/a_2 \approx .94 < 1 = max(1, .25) = max(1, a'_0/a_0)$.

[6] It is not always the case that an information triple in which one of the propositions is entailed by the two other propositions is more coherent than an information triple in

(iii) Information Sets of Unequal Size. Finally, we consider a comparison between an information pair and an information triple. The following example is inspired by the paradigmatic example of non-monotonic reasoning about Tweety the penguin. We are not interested in non-monotonic reasoning here, but merely in the question of the coherence of information sets. Suppose that we come to learn from independent sources that someone's pet Tweety is a bird (B) and that Tweety cannot fly, i.e. that Tweety is a ground-dweller (G). Considering what we know about pets, {B, G} is highly incoherent information. Aside from the occasional penguin, there are no ground-dwelling birds that qualify as pets, and aside from the occasional bat, there are no flying non-birds that qualify as pets. Later, we receive the new item of information that Tweety is a penguin (P). Our extended information set $S' = \{B, G, P\}$ seems to be much more coherent than $S = \{B, G\}$. So let us see whether our analysis bears out this intuition. We construct a joint probability distribution for B, G, and P together with the marginalized probability distributions for B and G in Figure 2.6.

Since the information sets are of unequal size, we need to appeal to our general method in (2.11) and construct the difference function:

$$(2.16) \quad f_{tweety} = f_r(S', S) = \frac{a'_0 + \bar{a}'_0\bar{r}^3}{a'_0 + a'_1\bar{r} + a'_2\bar{r}^2 + a'_3\bar{r}^3} - \frac{a_0 + \bar{a}_0\bar{r}^2}{a_0 + a_1\bar{r} + a_2\bar{r}^2}.$$

We have plotted f_{tweety} in Figure 2.7. This function is positive for all values of $r \in (0,1)$. We may conclude that S' is more coherent than S, which is precisely the intuition that we wanted to account for.

which the propositions are probabilistically independent. For instance, suppose that R_2 and R_3 are extremely incoherent propositions, i.e. the truth of R_2 makes R_3 extremely implausible and vice versa, and that R_1 is an extremely implausible proposition which in conjunction with R_2 entails R_3. Then it can be shown that this set of propositions is not a more coherent set than a set of probabilistically independent propositions. This is not unwelcome, since entailments by themselves should not warrant coherence. Certainly, $\{R_1, R_2, R_3\}$ should not be a coherent set when R_2 and R_3 are inconsistent and R_1 contradicts our background knowledge, although $R_1 \& R_2 \vdash R_3$. A judgement to the effect that S is more coherent than S' depends both on logical relationships and background knowledge.

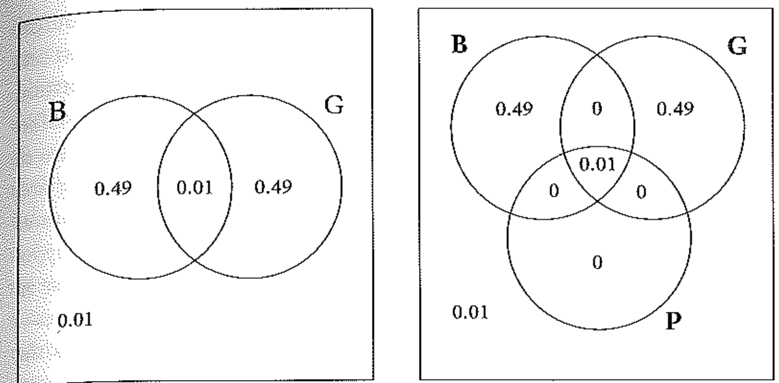

Fig. 2.6 A diagram for the probability distribution for Tweety before and after extension with [Tweety is a penguin]

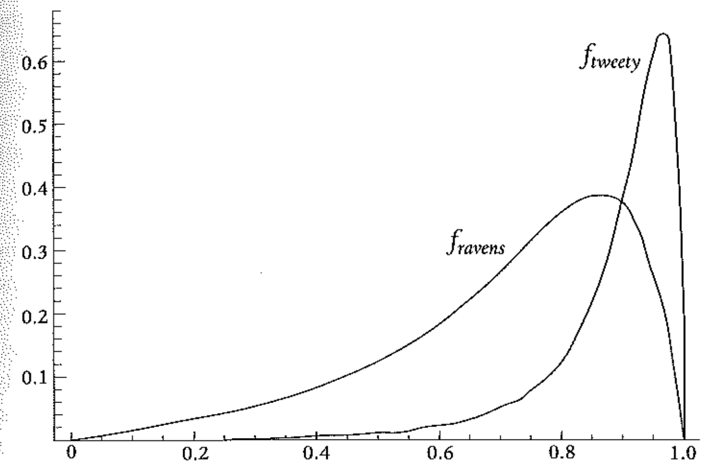

Fig. 2.7 The difference functions for BonJour's ravens and Tweety

2.4. EQUAL RELIABILITY

We have built into our model the assumption that the sources are equally reliable, i.e. that all sources have the same true positive rate p and the same false positive rate q. This seems like an unreasonably strong assumption, since, when we are gathering information in the actual world, we typically trust some sources less and some sources more. But our assessment of the relative coherence of information sets has nothing to do with how much we *actually* trust our information sources. As a matter of fact, we may assess the coherence of an information set without having any clue whatsoever who the sources are of the items in this information set or what their degrees of reliability are. An assessment of coherence requires a certain metric that features *hypothetical* sources with certain idealized characteristics. These hypothetical sources are not epistemically perfect, as is usually the case in idealizations. Rather, they are characterized by idealized *imperfections*—their partial reliability. Furthermore, our idealized sources possess the same degree of *internal* reliability and the same degree of *external* reliability. By internal reliability we mean that the sources for each item within an information set are equally reliable, and by external reliability we mean that the sources for each information set are equally reliable.

To see why internal reliability is required in our model, consider the following two information sets. Set S contains two equivalent propositions R_1 and R_2 and a third proposition R_3 that is highly negatively relevant with respect to R_1 and R_2. Set S' contains three propositions R'_1, R'_2, and R'_3 and every two propositions in S' are just short of being equivalent. One can specify the contents of such information sets such as to make S' intuitively more coherent than S. Our formal analysis will agree with this intuition. Now suppose that it turns out that the actual—i.e. the non-idealized—information sources for R_1, R'_1, R_2, and R'_2 are quite reliable and for R_3 and R'_3 are close to fully unreliable. We assign certain values to the reliability parameters to reflect this situation and calculate the proportional confidence boosts that actually result for both information sets. Plausible values can be picked for the relevant parameters so that the proportional confidence boost for S actually *exceeds* the proportional confidence boost for S'. This comes about because the actual information sources virtually bring nothing to the propositions R_3 and R'_3 and because R_1 and R_2 are indeed equivalent (and hence maximally coherent), whereas R'_1 and R'_2 are short of being equivalent (and hence less than maximally coherent). But what we want is an assessment of the relative coherence of $\{R_1, R_2, R_3\}$ and $\{R'_1, R'_2, R'_3\}$ and not of the relative coherence of $\{R_1, R_2\}$ and $\{R'_1, R'_2\}$. The appeal to ideal agents with the same degree of internal reliability in our metric is warranted by the fact that we want to compare the degree of coherence of complete information sets and not of some proper subsets of them.

Second, to see why *external* reliability is required in our model, consider some information set S which is not maximally coherent, but clearly more coherent than an information set S'. Any of our examples in Section 2.3 will do for this purpose. It is always possible to pick two values r and r' so that $c_{r'}(S') > c_r(S)$. To obtain such a result, we need only pick a value of r' in the neighbourhood of 0 or 1 and pick a less extreme value for r, since it is clear from (2.6) that for r' approaching 0 or 1, $c_{r'}(S')$ approaches 1. This is why coherence needs to be assessed relative to idealized sources that are taken to have the same degree of external reliability.

2.5. INDETERMINACY

Our analysis has some curious repercussions for the indeterminacy of comparative judgements of coherence. Consider the much debated problem among Bayesians of how to set the prior probabilities. We have chosen examples in which shared background knowledge (or ignorance) imposes constraints on what prior joint probability distributions are reasonable.[7] In the case of the corpse in Tokyo, one could well imagine coming to the table with no prior knowledge whatsoever about where an object is located in a grid with equal-sized squares. Then it seems reasonable to assume a uniform distribution over the squares in the grid. In the case of BonJour's ravens we modelled a certain lack of ornithological knowledge and let the joint probability

[7] Note that this is no more than a framework of presentation. Our approach is actually neutral when it comes to interpretations of probability. Following Gillies (2000), we favour a pluralistic view of interpretations of probability. The notion used in a certain context depends on the application in question. But, if one believes, as a more zealous personalist, that only the Kolmogorov axioms and Bayesian updating impose constraints on what constitute reasonable degrees of confidence, then there will be less room for rational argument and intersubjective agreement about the relative coherence of information sets. Or, if one believes, as an objectivist, that joint probability distributions can only be meaningful when there is the requisite objective ground, then there will be less occasion for comparative coherence judgements. None of this affects our project. The methodology for the assessment of the coherence of information sets remains the same, no matter what interpretation of probability one embraces.

distribution respect the logical entailment relation between the propositions in question. In the case of Tweety one could make use of frequency information about some population of pets that constitutes the appropriate reference class.

But often we find ourselves in situations without such reasonable constraints. What are we to do then? For instance, what is the probability that the butler was the murderer (B), given that the murder was committed with a kitchen knife (K), that the butler was having an affair with the victim's wife (A), and that the murderer was wearing a butler jacket (J)? Certainly the prior joint probability distributions over the propositional variables B, K, A, and J may reasonably vary widely for different Bayesian agents and there is little that we can point to in order to adjudicate in this matter. But to say that there is room for legitimate disagreement among Bayesian agents is not to say that anything goes. Certainly we will want the joint probability distributions to respect, among others things, the feature that $P(B|K, A, J) > P(B)$. Sometimes there are enough rational constraints on degrees of confidence to warrant agreement in comparative coherence judgements over information sets. And sometimes there are not. It is perfectly possible for two rational agents to have degrees of confidence that are so different that they are unable to reach agreement about comparative coherence judgements. This is one kind of indeterminacy. Rational argument cannot always bring sufficient precision to degrees of confidence to yield agreement on judgements of coherence.

But what our analysis shows is that this is not the only kind of indeterminacy. Two rational agents may have the same subjective joint probability distribution over the relevant propositional variables and still be unable to make a comparative judgement about two information sets. This is so for situations δ and ε in the case of the corpse in Tokyo. Although there is no question about what constitutes the proper joint probability distributions that are associated with the information sets in question, no comparative coherence judgement about S^{δ} and S^{ε} is possible. This is so because the proportional confidence boost for S^{δ} exceeds the proportional confidence boost for S^{ε} for some intervals of the reliability parameter, and vice versa for other intervals. If coherence is to be measured by the proportional confidence boost and if it is to be independent of the reliability of the witnesses, then there will not exist a coherence ordering for some pairs of information sets.

In short, indeterminacy about coherence may come about because rationality does not sufficiently constrain the relevant degrees of confidence. In this case, it is our epistemic predicament with respect to the content of the information set that is to blame. However, even when the probabilistic features of a pair of information sets are fully transparent, it may still fail to be the case that one information set is more coherent than (or equally coherent as) the other. Prima facie judgements can be made on both sides, but no judgement *tout court* is warranted. In this case, indeterminacy is not due to our epistemic predicament, but rather to the probabilistic features of the information sets.

2.6. ALTERNATIVE PROPOSALS

We return to the alternative proposals to construct a coherence ranking that were introduced in Chapter 1 and will show that these proposals yield counter-intuitive results. First, Lewis does not propose a measure that induces an ordering over information sets. Rather, he claims that coherent (or, in his words, congruent) information sets have the following property

(2.17) $P(R_i|R_1, \ldots, R_{i-1}, R_{i+1}, \ldots, R_n) > P(R_i)$ for all $i = 1, \ldots, n$.

But let us suppose that an information set contains n pairs of equivalent propositions, but that there is a relation of strong negative relevance (but not of inconsistency) between the propositions in each pair and all other propositions. In other words, $P(R_i, R_j) > P(R_i, R_j| R_1, \ldots, R_{i-1}, R_{i+1}, \ldots, R_{j-1}, R_{j+1}, \ldots, R_{2n}) \approx 0$ but not equal to 0, for each equivalent pair of propositions $\{R_i, R_j\}$. Then one would be hard-pressed to say that this information set is coherent. And yet, according to Lewis, this information set is coherent, because, assuming non-extreme marginal probabilities, $1 = P(R_i|R_1, \ldots, R_{i-1}, R_{i+1}, \ldots, R_n) > P(R_i)$ for all $i = 1, \ldots, 2n$.[8]

Second, Shogenji proposes that

$$S \succeq S' \text{ iff } m_s(S) = \frac{P(R_1, \ldots, R_m)}{\prod_{i=1}^{n} P(R_i)} \geq \frac{P(R'_1, \ldots, R'_n)}{\prod_{i=1}^{n} P(R'_i)} = m_s(S'). \quad (2.18)$$

[8] For an example, see Bovens and Olsson (2000: 688–9).

The following example shows that the Shogenji measure is counter-intuitive. Suppose that there are 1,000 equiprobable suspects for a crime with equal proportions of Africans, North Americans, South Americans, Europeans, and Asians. Now consider the information sets S = {R_1 = [The culprit is either an African, a North American, a South American, or a European], R_2 = [The culprit is not Asian]} and S′ = {R'_1 = [The culprit is an African], R'_2 = [The culprit is either Youssou (a particular African), Sulla (a particular South American), or Pierre (a particular European)]}. Since S contains propositions that pick out coextensive sets of suspects, whereas there is relatively little overlap between the propositions in S′, it seems reasonable to say that S is a more coherent set than S′. However, on the Shogenji measure, $m_s(S) = \frac{.8}{.8 \times .8} = 1.25 < 1.67 = \frac{.001}{.2 \times .003} = m_s(S')$. Our procedure, on the other hand, clearly matches the intuitive result in this case. The proportional confidence boost measure c_r is maximal for the maximally coherent information set S containing equivalent propositions. Hence, the difference function $f_r(S, S') = c_r(S) - c_r(S') > 0$ for all values of $r \in (0, 1)$ and so, by (2.11), S is more coherent than S′.

Third, Olsson tentatively proposes that

(2.19) $S \succeq S'$ iff

$$m_o(S) = \frac{P(R_1, \ldots, R_m)}{P(R_1 \vee \ldots \vee R_m)} \geq \frac{P(R'_1, \ldots, R'_n)}{P(R'_1 \vee \ldots \vee R'_n)} = m_o(S').$$

The Tweety example shows that this measure is counter-intuitive. It seems reasonable to say that the information *pair* S = {[My pet Tweety is a bird], [My pet Tweety cannot fly]} is less coherent than the information *triple* S′ = {[My pet Tweety is a bird], [My pet Tweety cannot fly], [My pet Tweety is a penguin]}. But from Figure 2.6 we can read off that $m_o(S) = .01/.99 = m_o(S')$.

Fourth, we focus on Fitelson's measure as applied to information *pairs*. The Kemeny–Oppenheim measure is a measure of factual support when the marginal probabilities of R_1 and R_2 are not extreme:

(2.20) $$F(R_1, R_2) = \frac{P(R_1|R_2) - P(R_1|\neg R_2)}{P(R_1|R_2) + P(R_1|\neg R_2)}$$
for $P(R_1) < 1$ and $P(R_2) > 0$.

Fitelson proposes that

(2.21) $S \succeq S'$ iff

$$m_f(S) = \frac{F(R_1, R_2) + F(R_2, R_1)}{2} \geq \frac{F(R'_1, R'_2) + F(R'_2, R'_1)}{2} = m_f(S').$$

The following example shows that this measure yields counter-intuitive results. Let there be 100 suspects for a crime who have an equal chance of being the culprit. In situation one, let there be 6 Trobriand suspects and 6 chess-playing suspects; there is 1 Trobriand chess player. In situation two, let there be 85 Ik suspects and 85 rugby-playing suspects; there are 80 Ik rugby players. Which information is more coherent—S = {R_1 = [The culprit is a Trobriand], R_2 = [The culprit is a chess player]} or S′ = {R'_1 = [The culprit is an Ik], R'_2 = [The culprit is a rugby player]}? The information in S′ seems to fit together much better than in S, since there is so little overlap between being a Trobriander and being a chess player and there is considerable overlap between being an Ik and a rugby player. But note that on Fitelson's measure $m_f(S) \approx .52 > .48 \approx m_f(S')$. The Fitelson measure behaves curiously for cases in which we increase the overlapping area, while keeping the non-overlapping area fixed. Intuitively, one would think that when keeping the non-overlapping area fixed, then, the more overlap, the greater the coherence. And this is indeed what our condition (2.12) indicates. But on the Fitelson measure, this is not the case. In Figure 2.8, we set the non-overlapping area at $P(R_1, \neg R_2) = P(\neg R_1, R_2) = .05$. We increase the overlapping area a_0 from .01 to .80 and plot the Fitelson measure as a function of a_0 in Figure 2.9. The measure first increases from $a_0 = .01$ and then reaches its maximum for $a_0 \approx .17$ and subsequently decreases again. We fail to see any intuitive justification for this behaviour of the measure.

Where do these proposals go wrong? Lewis forgets that strong positive relevance between each proposition in a singleton set and the propositions in the complementary set is compatible with strong negative relevance between certain propositions in the information set. On Shogenji's measure, information sets containing less probable propositions tend to do better on the coherence score, so much so that information sets with non-equivalent but less probable propositions may

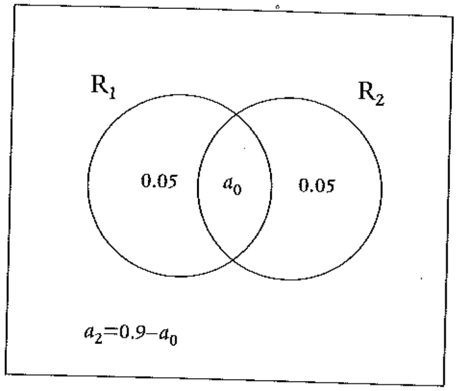

Fig. 2.8 A diagram for the probability distributions of the information sets in our counter-example to the Fitelson measure

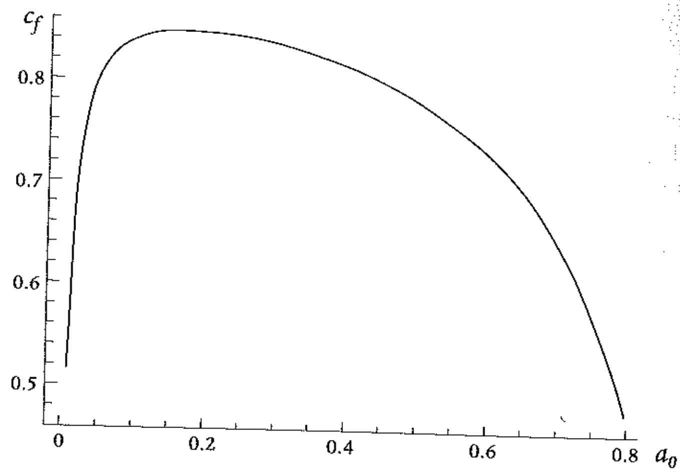

Fig. 2.9 The Fitelson measure m_f as a function of $a_0 \in [.01, .8]$ for the information sets in our counter-example to the Fitelson measure

score higher than information sets containing all and only equivalent propositions. We concur with Fitelson (2003) that an information set with all and only equivalent propositions is maximally coherent. It is not possible for non-equivalent propositions to fit together better than equivalent propositions. Certainly, information sets with less probable propositions may be more informative—it is more informative when a suspect points to Sulla than when she points to the

whole group of South Americans. Furthermore, informativeness is a good-making characteristic of witness reports, as is coherence. But this is no reason to think that informativeness should be an aspect of coherence. Olsson pays exclusive attention to the relative overlap between the propositions in the information set. But note that by increasing the number of propositions one can increase relations of positive relevance while keeping the relative overlap fixed. Fitelson's measure assesses the degree of positive relevance between the propositions in the information set. But sometimes the relative overlap between the propositions gets the upper hand in our intuitive judgement of coherence.

We believe that judgements of coherence rest on the subtle interplay between the degree of positive relevance relations and relative overlap relations between propositions. To determine the nature of this subtle interplay, it is of no use to consult our intuitions. Rather, one needs to determine the relative coherence through the role that coherence is meant to play—the role of boosting our confidence in the propositions in question. More coherent information sets are information sets that display higher proportional coherence boosts regardless of the degree of reliability of the sources.

2.7. THEORY CHOICE IN SCIENCE

Where does our analysis leave the claim in philosophy of science that coherence plays a role in theory choice? We repeat the equality in (2.8):

$$P^*(R_1, \ldots, R_n) = \frac{a_0}{a_0 + \bar{a}_0 \bar{r}^n} \times c_r(\{R_1, \ldots, R_n\}). \tag{2.22}$$

What this means is that our degree of confidence in an information set S can be expressed in terms of the measure $c_r(S)$ which induces a quasi-ordering weighted by a factor. Note that this factor approximates 1 for larger information sets (large n) as well as for highly reliable sources ($r \approx 1$). Let us assume that we are comparing two information sets that can be ordered. Then the relative degree of confidence for these two information sets is fully determined by their relative coherence, if either the sources are sufficiently reliable or the information sets are sufficiently large.

One can represent a scientific theory T by a set of propositions $\{T_1, \ldots, T_m\}$. Let the T_is be assumptions, scientific laws, specifications of parameters, and so on. It is not plausible to claim that each proposition is independently tested, i.e. that each T_i shields off the evidence E_i for this proposition from all other propositions in the theory and all other evidence. The constitutive propositions of a theory are tested in unison. They are arranged into models that combine various propositions in the theory. Different models typically share some of their contents, i.e. some propositions in T may play a role in multiple models. It is more plausible to claim that each model M_i is being supported by some set of evidence E_i and that each M_i shields off the evidence E_i in support of the model from the other models in the theory and from other evidence. This is what it means for the models to be supported by independent evidence. There are complex probabilistic relations between the various models in the theory.

Formally, let each M_i for $i = 1, \ldots, n$ combine the relevant propositions of a theory T that are necessary to account for the independent evidence E_i. A theory T can be represented as the union of these M_is.[9] Let M_i be the variable which ranges over the value M_i stating that all propositions in the model are true and the value $\neg M_i$ stating that at least one proposition in the model is false. In Bayesian confirmation theory, E_i is evidence for M_i if and only if the likelihood ratio

$$x_i = \frac{P(E_i|\neg M_i)}{P(E_i|M_i)} \tag{2.23}$$

is contained in (0,1). Hence, E_i stands to M_i in the same way as $REPR_i$ stands to R_i in our framework. Let us suppose that all the likelihood ratios x_i equal x. $\bar{x} := 1 - x$ now plays the same role as r in our earlier model. We can construct a probability measure P for the constituent models of a theory T and identify the weight vector $< a_0, \ldots, a_n >$. If we translate the constraints of our earlier model, the following result holds up:

[9] This account of what a scientific theory is contains elements of both the syntactic view and the semantic view. Scientific theories are characterized by the set of their models, as on the semantic view, and these models (as well as the evidence for the models) are expressed as sets of propositions, as on the syntactic view.

$$P^*(M_1, \ldots, M_n) = \frac{a_0}{a_0 + \bar{a}_0 x^n} \times c_{\bar{x}}(\{M_1, \ldots, M_n\}). \tag{2.24}$$

Suppose that we are faced with two contending theories. The models within each theory are supported by independent items of evidence. It follows from (2.24) that, if (i) the evidence for each model is equally strong, as expressed by a single parameter x, *and*, (ii) either the evidence for each model is relatively strong ($x \approx 0$), *or*, each theory can be represented by a sufficiently large set of models (large n), then a higher degree of confidence is warranted for the theory that is represented by the more coherent set of models. Of course, we should not forget the *caveat* that indeterminacy springs from two sources. First, there may be substantial disagreement about the prior joint probability distribution over the variables $M_1, \ldots, M_n$, and second, even in the absence of such disagreement, no comparative coherence judgement may be possible between both theories, represented by their respective constitutive models. But even in the face of our assumptions and the *caveats* concerning indeterminacy, this is certainly not a trivial result about the role of coherence in theory choice within the framework of Bayesian confirmation theory.