Coherentism, Reliability and Bayesian Networks

LUC BOVENS

ERIK J. OLSSON

The coherentist theory of justification provides a response to the sceptical challenge: even though the independent processes by which we gather information about the world may be of dubious quality, the internal coherence of the information provides the justification for our empirical beliefs. This central canon of the coherence theory of justification is tested within the framework of Bayesian networks, which is a theory of probabilistic reasoning in artificial intelligence. We interpret the independence of the information gathering processes (IGPs) in terms of conditional independences, construct a minimal sufficient condition for a coherence ranking of information sets and assess whether the confidence boost that results from receiving information through independent IGPs is indeed a positive function of the coherence of the information set. There are multiple interpretations of what constitute IGPs of dubious quality. Do we know our IGPs to be no better than randomization processes? Or, do we know them to be better than randomization processes but not quite fully reliable, and if so, what is the nature of this lack of full reliability? Or, do we not know whether they are fully reliable or not? Within the latter interpretation, does learning something about the quality of some IGPs teach us anything about the quality of the other IGPs? The Bayesian-network models demonstrate that the success of the coherentist canon is contingent on what interpretation one endorses of the claim that our IGPs are of dubious quality.

What is wrong with my mulberry tree? Will there be a thunderstorm tomorrow? Who is spreading nasty rumours about me in town? To find out such matters, we ask our friends, go and take a close look ourselves, run some tests … whatever seems to do the job. Once we have gathered some information that is more or less reliable through various channels, we step back and assess what the story looks like so far. Does it make any sense? Does it hang together in some fashion or other? All other things being equal, we are more inclined to believe a story that hangs together than a story that does not. “Got it!” we exclaim, when things fall in place like the pieces of a jigsaw puzzle. “Yeah right—likely story!” we mumble to ourselves, when the truth of part of our story would make other parts extremely unlikely.

Mind, Vol. 109 . 436 . October 2000

So far, there does not seem to be much to quibble about. It is all just common sense, so they say. Certainly, whether some story or other is credible has something to do with how well the various parts of the story support one another. But it is this innocent item of common sense that has given rise to the much debated coherence theory of justification in epistemology. The central idea in this theory is that whether a proposition is likely to be true or not is at least partially determined by how well it coheres with the other propositions that are also candidates for belief. The route from common sense to philosophical theory requires a precise account of what common sense has to offer as well as a critical examination of the conditions under which this account is correct.

C. I. Lewis (1946, p. 346) takes a case of "relatively unreliable witnesses who independently tell the same circumstantial story" to be a paradigm case of the justificatory role of coherence. This is just a limiting case of coherence. The witnesses need not provide us with precisely the same story, but rather, it suffices that their respective stories accord with one another:

> For any one of these reports taken singly, the extent to which it confirms what is reported may be slight. And antecedently, the probability of what is reported may also be small. But congruence [i.e. coherence] of the reports establishes a high probability of what they agree upon.

Laurence Bonjour remarks:

> What Lewis does not see, however, is that his own example shows quite convincingly that no antecedent degree of warrant or credibility is required. For as long as we are confident that the reports of various witnesses are genuinely independent of each other, a high enough degree of coherence among them will eventually dictate the hypothesis of truth telling as the only available explanation of their agreement—even, indeed, if those individual reports initially have a high degree of negative credibility, that is, are much more likely to be false than true (for example in the case where all of the witnesses are known to be habitual liars). (Bonjour 1985, p. 148)

The same can be said when we gain information through direct observation rather than through witnesses. In Bonjour's terminology, direct observation provides us with *cognitively spontaneous beliefs*. It is the coherence of these cognitively spontaneous beliefs that renders them credible: "… so long as apparently cognitively spontaneous beliefs are genuinely independent of each other, their agreement will generate credibility, without the need for an initial degree of warrant" (Bonjour 1985, p. 148).

The ingredients of the coherentist canon are all on the table. We gather information through many types of processes: we consult witnesses, make direct observations… Let an information set be a set of propositions acquired through such processes. Whether the content of an information set is credible is what is in question. Now suppose (i) that the reports are independent, (ii) that antecedently, the content of the information set is less than fully credible and (iii) that the information gathering processes are dubious. Then, says the coherentist, the more coherent the information set is, the more credible its content will become once all the reports are in.

We shall represent the credibility of the content of an information set in a probabilistic framework: it is measured by the joint probability of the propositions in the information set. The antecedent credibility is measured by the prior joint probability of the propositions in the information set, that is, the joint probability before any of the reports have come in. The credibility after all the reports are in is measured by the posterior joint probability of the propositions in the information set, that is, the joint probability of the propositions in the information set, conditional on all the reports being in.

We are going to take the coherentist to task. The background conditions (i) and (ii) need some interpretation but can be made precise. However, there are multiple interpretations of background condition (iii): the coherentist canon is true on some interpretations of "dubious information gathering processes", while it is not on other interpretations. We will construct a series of interpretations against the backdrop of theoretical work on conditional independence and test the coherentist canon by means of Bayesian networks. The result will read like a flow chart that determines on what interpretations the coherentist canon is correct and on what interpretations it is not.

1. Coherence

What is coherence? It is perhaps the most persistent complaint against coherence theories that a definition of this central notion is not forthcoming. Already in 1934, A.C. Ewing writes that the absence of such a definition reduces the theory "to the mere uttering of a word, coherence, which can be interpreted so as to cover all arguments, but only by making its meaning so wide as to rob it of almost all significance" (Ewing 1934, p. 246). Bonjour's *The Structure of Empirical Knowledge* is one of the standard contemporary defences of a coherence theory. Marshall Swain

(Swain 1989, p. 116) writes that "[o]ne of the most disappointing features of Bonjour's book is the lack of detail provided in connection with the central notion of coherence". Bonjour (1985, p. 101) himself admits that his account is "a long way from being as definitive as desirable". Most recently he writes that "the precise nature of coherence remains an unsolved problem" (1999, p. 124) and

> spelling out the details of this idea, particularly in a way that would allow reasonably precise assessments of comparative coherence, is extremely difficult, at least partly because such an account will depend on the correct account of a number of more specific and still inadequately understood topics, such as induction, confirmation, probability, explanation and various issues in logic … (1999, p. 123).

What coherentists do agree on is that a coherent information set is an information set whose members provide mutual support to each other. This suggests the comparative notion that an information set is the more coherent, the more mutual support its members provide to each other. To avoid computational complexity, we will focus on information *pairs* and will spell out a minimal sufficient condition for the relation "… more coherent than …" in probabilistic terms. An information pair is the more coherent, the more likely each proposition becomes given the truth of the other proposition:

(*Coherence*) For an information pair {A,B} and probability distributions P and P', if $P(A|B) > P'(A|B)$ and $P(B|A) > P'(B|A)$, then {A,B} is more coherent on probability distribution P than on probability distribution P'.[1]

This condition can be extended to information sets that contain more than two elements, but one should not be too rash: it does not suffice to consider the conditional probability of each proposition in the information set on the remaining propositions. Consider a population containing a reasonable number of students and a reasonable number of octogenarians. Suppose that all and only students like to party and that all and only octogenarians are bird watchers, and that there are some, but very few

[1] Even this minimal condition is not entirely innocent. Let there be a roulette wheel with one hundred numbers and an equal chance for each number to be the winning number (WN). The wheel is spun. On scenario a, Joe says that WN is 49 or 50 and Amy says that WN is 50 or 51. On scenario b, Joe says that WN is 1, 2, …, or 70 and Amy says that WN is 31, 32, …, 100. Then, on our minimal condition, the information pair S = {[Joe is correct], [Amy is correct]} is more coherent on P^b than on P^a, which seems counterintuitive. (Following Quine (1960, p. 168), the square brackets are used to refer to the proposition expressed by the enclosed sentence.) There are two responses. One could argue that *coherence* is an ambiguous notion. One can think of coherence as a measure of *agreement* or

octogenarian students. A murder happened in town and we receive the following reports from some dubious sources: (i) the suspect is a student and (ii) the suspect is an octogenarian. Clearly this information set would score low on coherence: both the probability that the suspect is a student given that he is an octogenarian, and the probability that the suspect is an octogenarian given that he is a student, are low. But now suppose that we receive the following additional information: (iii) the suspect likes to party and (iv) the suspect likes to watch birds. We would be hard-pressed to call this information set coherent: one half of the story is highly unlikely given the other half of the story. But if we only consider the conditional probability of each element on the remaining elements, the information set would receive a maximal score on coherence: the probability that the suspect is a student given that he likes to party, is an octogenarian and likes to watch birds is one, and the same holds true for the conditional probability of each proposition in the information set on the remaining propositions. The moral is clear: the degree of coherence of an information set is not only determined by the conditional probabilities of each single proposition on the remaining propositions in the set; rather, one should consider each genuine subset of the information set and assess the conditional probability of the propositions in this subset on the remaining propositions in the set. A minimal sufficient condition could then be spelled out for the coherence of information sets in general: if it is the case that these conditional probabilities are all greater on P than on P', then the information set is more coherent on P than on P'. The computational complexity grows exponentially with the number of propositions in the information set. We will focus on information pairs, since nothing is lost conceptually by this restriction.

Our condition yields a partial ordering. Let $P(A|B) > P'(A|B)$ and $P'(B|A) > P(B|A)$ for some information pair: in such cases, our condition has nothing to offer. We are out to test under what conditions the relative coherence of an information set affects its credibility. Why should we not restrict ourselves to information sets whose relative coherence we

as a measure of *striking agreement*. Coherence is sensitive to the specificity of the information on the latter but not on the former notion. The result is only counterintuitive on the latter notion, but we present a minimal condition for the former notion. This distinction is discussed in Olsson (2000). Alternatively, one could weaken the minimal condition even further by conjoining "... and $P(A,B) = P'(A,B)$" to the antecedent of the conditional. This avoids the counter example and makes no difference to our results, since we stipulate in § 2.2 that the prior joint probabilities should be equal if we want to assess the influence of the coherence of the information set on the posterior joint probability. This response is consistent with the criterion for a partial coherence ordering in Bovens and Hartmann (2000a and 2000b).

can confidently assess? Nothing would be gained by bringing in information sets whose relative coherence is in dispute. This is also sound scientific practice: if we wish to study, say, the effects of stress on life expectancy under certain dietary conditions, then we consider subjects whose relative stress levels we can confidently assess. It would be no objection to our study that our condition of relative stress cannot determine which of any two persons leads the more stressful life.

2. The Background Conditions

2.1 The independence of the reports.

This condition is a simplifying assumption. What we are interested in is whether the credibility of the content of an information set is affected by its coherence on various interpretations of "dubious information gathering processes". The simplest background condition to investigate this relation is that the reports are influenced only by the respective facts that they report on: they are not influenced by what other reporters have to say, nor are they influenced by other facts than the facts they report on. Sometimes this condition is fulfilled in the real world, sometimes it is not. One could investigate how various types and degrees of dependence of reports affect the relation between coherence and credibility for particular interpretations of "dubious gathering information processes", though this project will have to await another occasion.

How can we translate this condition into a probabilistic framework? Suppose that some item in the information set is true. Then, depending on how reliable the information gathering process is, there will be a certain chance that there will be a report to this effect. Similarly, if this item in the information set is false. But now suppose in addition that a report was (or was not) received concerning some other proposition in the information set; or suppose that some other proposition in the information set is true (or false). Then what our background condition stipulates is that such additional suppositions should not affect the chances that a report of the first item of information will be received, conditional on this item being true or conditional on this item being false. Information gathering processes tune their ears at best directly to the facts of the world that they report about: they may not always hear things correctly, but never ever do they tune their ears directly to what other information gathering processes have to report or to other facts of the world.

2.2. Antecedently, the content of the information set is less than fully credible.

The central idea of the coherentist canon is that more coherent information sets do better at raising the posterior joint probability of the propositions in the information set over and above the prior joint probability, even if the information gathering processes are dubious. Hence, we cannot set the prior joint probability at 1. Neither can we set it at 0 for that matter, since no amount of information could change such an assignment. For dramatic purposes, it is fitting to pick a relatively low prior joint probability and to let the coherence of the information set raise the posterior joint probability to a respectable level. But any value between both extremes will do.

What *is* crucial nonetheless, is that in assessing the influence of the coherence of the information set on the posterior joint probability of the propositions in the information set, we keep the prior joint probability of the propositions in the information set fixed. Otherwise it would be all too easy to disprove the coherentist canon. Suppose that two independent reporters provide us with the same very improbable reports, say that a dog died due to witchcraft. Suppose furthermore that two independent reporters provide us with reports that are different but not mutually exclusive, say that the dog died due to cancer and that the dog died due to pneumonia. We had no prior reason to trust the former pair of reporters any more or less than the latter pair. Clearly the former reports are more coherent than the latter reports. And yet, if we are somewhat sane, it is to be expected that the posterior (joint) probability that the dog died due to witchcraft is still lower than that the dog died due to the combined effect of cancer and pneumonia. The reason is clear: the prior probability of the dog dying due to witchcraft is much lower than the prior probability of the dog dying due to the combined effect of cancer and pneumonia. Once again, this is just sound scientific practice: if we wish to investigate the effect of the coherence of the information set on the posterior joint probability of the propositions in the information set, then we need to keep all other factors that may affect the posterior joint probability fixed; and of course the prior probability of the propositions in the information set is precisely such a factor.

2.3 The information gathering processes are dubious.

There are multiple interpretations of this background condition. The first three conditions are variations of the stipulation that we believe that our information gathering processes are *not* reliable:

(i) We believe our information gathering processes to be no better than randomizing processes: listening to their reports is not any more informative than flipping a coin, casting a die, spinning a roulette wheel...

(ii) We believe our information gathering processes to be halfway reliable: information gathering processes do not tell us that something is the case if and only if it is indeed the case, but they are more informative than randomizing processes. The chance that they tell us that something is the case given that it is the case, is greater than given that it is not the case. We also expect that they do occasionally yield false positives, that is, the chance that they tell us that something is the case, given that it is not the case, is real.

(iii) We believe our information gathering processes to be halfway reliable, but we do trust that they never yield false positives: sometimes they report that something is not the case given that it is the case, but never do they report that something is the case given that it is not the case.[2]

The next two conditions are variations of the stipulation that we do *not* believe that our information gathering processes are reliable. Suppose that we believe that there is some chance that they are (fully) reliable and that there is some chance that they are no better than randomizing processes. Then the following distinction will turn out to be relevant:

(iv) We do not believe that our information gathering processes are reliable, but if we would come to learn that all our information gathering processes but one are reliable or that all but one are no better than randomizing processes, this would not in the least affect our beliefs about whether the one remaining information gathering process is reliable or is no better than a randomizing process.

(v) We do not believe that our information gathering processes are reliable, but if we would come to learn that one of our information gathering processes is reliable, then we would believe that all of our information gathering processes are reliable. Similarly, if we

[2] This case is not a theoretical artifact, since information gathering processes often operate under the following conditions. We are testing a hypothesis which predicts that a particular variable takes on one out of a large range of possible values. Now suppose that we do believe that the information gathering process which assesses the variable in question is accurate some of the time and is no better than a randomizing process at other times. If this so, then we may expect false negatives: when the process is in randomizing mode, it is very likely that it will report that the predicted value of the variable does not hold given that it does hold. But false positives would be very unlikely: even if the process is in randomizing mode, it is extremely unlikely that it will report that the predicted value does hold given that it does not hold. Case (iii) is a limiting case of these conditions.

> would come to learn that one of our information gathering processes is no better than a randomizing process, we would believe that all of our information gathering processes are no better than randomizing processes.

This distinction can also be expressed as an (in)dependence condition which is distinct from the independence condition for reports. In (iv), the quality of our information gathering processes is independent: even if all but one are reliable, this tells us nothing about the last one; even if all but one are no better than randomizing processes, this tells us nothing about the last one. In (v), the quality of our information gathering processes is maximimally dependent: if one is reliable, all of them are; if one is no better than a randomizing process, all of them are. This distinction is relevant in two respects.

First, sets of information gathering processes can be more or less diverse in kind. For heterogeneous sets of information gathering processes, something in the direction of interpretation (iv) is more accurate: if we come to learn that one or more of our senses are reliable, then it is not clear that this should increase our confidence in our other senses, in our memory, in witnesses… For homogeneous sets of information processes, something in the direction of interpretation (v) is more accurate: if we come to learn that we are good at recognizing one colour, then it is more likely that we are good at recognizing other colours as well.

Second, the issue also touches on scepticism about induction. The extreme anti-inductivist claims that, just as millennia of sunrises cannot increase his confidence that the sun will rise tomorrow, coming to learn that all information gathering processes but one are reliable does not change his view about the reliability of that last information gathering process. He should cherish interpretation (iv). On the other hand, the extreme inductivist claims that, just as seeing one white swan suffices to conclude that all swans are white, coming to learn that one information gathering process is reliable suffices to conclude that all are reliable. She should cherish interpretation (v).

Real sets of information gathering processes are somewhere in between being strictly homogeneous and strictly heterogeneous. Doxastic sanity lies somewhere between extreme anti-inductivism and extreme inductivism. It will be easy to represent a continuum between these interpretations in our model.

We are now in a position to state the coherentist canon in a testable format. Let an information pair contain two propositions and let there be two probability distributions P and P' over the corresponding propositional variables. We have received reports that these propositions are true. If (i) the reports are independent, (ii) the prior joint probability of the propositions in the information pair is between 0 and 1 and (iii) a determinate interpretation of the claim that the information gathering processes are dubious is agreed upon, then the following thesis holds: if the information pair is more coherent on probability distribution P than on P', then the posterior joint probability of the propositions in the information pair is greater on distribution P than on distribution P'.

3. A Working Example

Let us return to our sorry-looking mulberry tree. Suppose that we have received independent reports that our tree has aphids and that our tree has fungus. We may have done tests, contacted a fungus specialist and an aphid specialist, or just taken a good look and noticed some critters and blotches that seem to match the pictures in our gardening book. The information pair contains the propositions that the tree is afflicted by aphids and that the tree is afflicted by fungus.

Our next task is to vary the coherence of this information pair. Suppose that pests can be classified according to their social behaviour. There are reclusive pests. They tend to crowd each other out: the presence of one type makes it less probable to find another type. There are sociable pests. They tend to invite in other pests: the presence of one type makes it more probable to find another type. There are indifferent pests. They do not tend to invite other pests in nor to crowd each other out: the presence of one type does not affect the probability of finding another type. We distinguish between three scenarios: (i) aphids and fungus are reclusive pests; (ii) aphids and fungus are indifferent pests and (iii) aphids and fungus are sociable pests. On the reclusive-pests scenario, the supposition that the tree has aphids lowers the prior probability that the tree has fungus:

(1) $P^r(\text{fungus}|\text{aphids}) < P^r(\text{fungus})$.

It follows by the probability calculus that:

$$P^r(\text{fungus}) < P^r(\text{fungus}|\text{not-aphids}).$$

In the indifferent-pests scenario, the supposition that the tree has aphids does not change the prior probability that the tree has fungus:

(2) $P^i(\text{fungus}|\text{aphids}) = P^i(\text{fungus})$

and hence:

$$P^i(\text{fungus}) = P^i(\text{fungus}|\text{non-aphids}).$$

In the sociable-pests scenario, the supposition that the tree has aphids raises the prior probability that the tree has fungus:

(3) $P^s(\text{fungus}|\text{aphids}) > P^s(\text{fungus})$

and hence:

$P^s\,(\text{fungus}) > P^s\,(\text{fungus}|\text{not-aphids})$.

To keep things simple, let us have fungus and aphids mirror image each other within each scenario. With P ranging over P^r, P^i and P^s,

(4) $P(\text{fungus}) = P(\text{aphids})$.

We fix the prior joint probability that the tree is afflicted by aphids and that the tree is afflicted by fungus at a determinate value between 0 and 1:

(5) $0 < P^r(\text{aphids, fungus}) = P^i(\text{aphids, fungus}) = P^s(\text{aphids, fungus}) < 1$.

It follows from (1) through (5) by the probability calculus that:

(6) $P^r(\text{fungus}|\text{aphids}) < P^i(\text{fungus}|\text{aphids}) < P^s(\text{fungus}|\text{aphids})$;

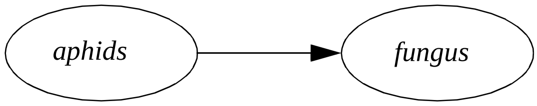
(7) $P^r(\text{aphids}|\text{fungus}) < P^i(\text{aphids}|\text{fungus}) < P^s(\text{aphids}|\text{fungus})$.

Hence, the information pair is more coherent on the probability distribution P^s than on P^i and is more coherent on P^i than on P^r.

Let us agree on some definite numbers and embed the probabilistic information in Bayesian networks. A Bayesian network allows for an economical representation of a joint probability distribution over a set of variables. It organizes the variables into a *Directed Acyclical Graph* (DAG) which encodes (conditional) independences. A DAG is a set of *nodes* and a set of *arrows* between these nodes under the constraint that one does not run into a cycle by following the direction of the nodes. Each node represents a variable. The node at the tail of an arrow is the *parent node* of the node at the head, and the node at the head is the *child node* of the node at the tail. A Bayesian network contains a probability distribution for the variable in each *root* node (i.e. in each unparented node), and a probability distribution for the variable in each child node, conditional on any combination of values of the variables in their parent nodes. When implemented on a computer, a Bayesian network performs complex probabilistic calculations with one keystroke.

The variables in our example are the propositional variables *aphids*, *fungus*, *infoaphids* and *infofungus*. The propositional variable *aphids* can take on two values, viz. aphids (i.e. the proposition that the tree has aphids) and non-aphids (i.e. the proposition that the tree does not have aphids). Similarly for the propositional variable *fungus*. We represent *aphids* by a root node and *fungus* by a child node. Then we assign a value to $P(\text{aphids})$ and values to $P(\text{fungus}|\text{aphids})$ and $P(\text{fungus}|\text{not-aphids})$ in each scenario. The following assignments respect (1) through

(7) and the prior joint probability of aphids and fungus is .25 in each scenario.[3]



Reclusive pests:

$P^r(\text{aphids}) = \frac{5}{8}$ $\quad$ $P^r(\text{fungus} \mid \text{aphids}) = \frac{2}{5}$

$P^r(\text{fungus} \mid \text{non-aphids}) = 1$

Indifferent pests:

$P^i(\text{aphids}) = \frac{1}{2}$ $\quad$ $P^i(\text{fungus} \mid \text{aphids}) = \frac{1}{2}$

$P^i(\text{fungus} \mid \text{non-aphids}) = \frac{1}{2}$

Sociable pests:

$P^s(\text{aphids}) = \frac{1}{3}$ $\quad$ $P^s(\text{fungus} \mid \text{aphids}) = \frac{3}{4}$

$P^s(\text{fungus} \mid \text{non-aphids}) = \frac{1}{8}$

$P^r(\text{aphids, fungus}) = P^i(\text{aphids, fungus}) = P^s(\text{aphids, fungus}) = \frac{1}{4}$

Figure 1: The social behaviour of pests

We must make sure that the reports are independent. This means that the chance that we will receive the information that the tree has aphids is fully determined by whether the tree has aphids or not and by nothing else. The additional suppositions that we do (or do not) receive the information that the tree has fungus or that the tree actually has (or does not have) fungus are irrelevant. Let the variable *infoaphids* take on two values, viz. infoaphids (i.e. the proposition that we receive a report that our tree has aphids upon consultation with an aphids specialist) and non-infoaphids (i.e. the proposition that we do not receive a report that the tree has aphids upon consultation with an aphids specialist). Similarly

[3] There is actually no need for an arrow in a Bayesian Network when aphids and fungus are indifferent pests, since the variables *aphids* and *fungus* are independent in this scenario. But since we are interested in comparing degrees of coherence, it is more perspicuous to retain the arrow in this scenario as well.

for *infofungus*. The reports are independent just in case for all combinations of values of *aphids*, *fungus* and *infofungus*, $P(\text{infoaphids}|\textit{aphids}) = P(\text{infoaphids}|\textit{aphids}, \textit{fungus}, \textit{infofungus})$. The following shorthand notation is customary:

(8) $\textit{infoaphids} \perp\!\!\!\perp \textit{fungus}, \textit{infofungus} \mid \textit{aphids}$

or, in words, the variable *infoaphids* is independent of the variables *fungus* and *infofungus*, conditional on the variable *aphids*. Similarly,

(9) $\textit{infofungus} \perp\!\!\!\perp \textit{aphids}, \textit{infoaphids} \mid \textit{fungus}$.

To represent this information in our network, we add two nodes to the network, viz. *infoaphids* and *infofungus*, and arrows from *aphids* to *infoaphids* and from *fungus* to *infofungus*.

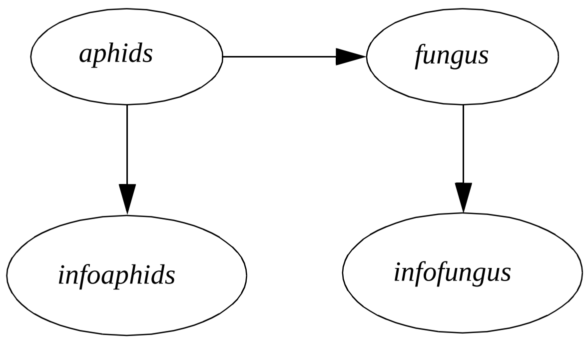


Figure 2: Independent reports

There is a certain heuristic that underlies the direction of these arrows: in the causal order of the world, the presence or absence of aphids or fungus has a *direct influence* on whether the information that the tree has aphids or fungus gains access to our information set. But this is only a heuristic. The Parental Markov Condition (Pearl 2000: 19) states the precise probabilistic meaning of the arrows in the DAG:

(*PMC*) Each variable in the network is independent of its non-descendants, conditional on all of its parents.

Infoaphids is a child to the single-parent *aphids*. Neither *fungus* nor *infofungus* are descendants of *infoaphids*. Hence, what the arrow from *aphids* to *infoaphids* expresses is that *infoaphids* is independent of *fungus* and *infofungus*, conditional on *aphids*, that is, condition (8). Similarly, we can read off condition (9) of the network.[4]

We will now turn to the respective interpretations of "dubious information gathering processes". Interpretations (i) through (iii) can be modelled in our network and will be presented in section 5. Interpretations (iv) and (v) require an extension of the network and will be presented in section 6.

4. We Believe That Our Information Gathering Processes Are Not Reliable

To keep matters simple, let us stipulate that the information gathering processes for aphids and fungus are equally informative:

(10) $P(\text{infoaphids}|\text{aphids}) = P(\text{infofungus}|\text{fungus})$

and

(11) $P(\text{infoaphids}|\text{non-aphids}) = P(\text{infofungus}|\text{non-fungus})$.

On interpretation (i) the information gathering processes are no better than randomizing processes:

(12) $1 > P(\text{infoaphids}|\text{aphids}) = P(\text{infoaphids}|\text{non-aphids}) > 0$.

On interpretation (ii) the information gathering processes are halfway reliable and false positives are not excluded:

(13) $P(\text{infoaphids}|\text{aphids}) > P(\text{infoaphids}|\text{non-aphids}) > 0$.

On interpretation (iii) the information gathering processes are halfway reliable but false positives are excluded:

(14) $1 > P(\text{infoaphids}|\text{aphids}) > P(\text{infoaphids}|\text{non-aphids}) = 0$

In figure 3 we have picked some definite values for the link from *aphids* to *infoaphids* for each interpretation. The values are uniform across the scenarios of reclusive, indifferent and sociable pests.

[4] Why could we be so cavalier about the direction of the arrows between *aphids* and *fungus*? When it comes to *aphids* and *fungus*, we have no clue about the causal order: do the aphids attract or deter the fungus? Or vice versa? Or is there a complex pattern of causal interactions? We do not know and the heuristic is of no consequence in this case. A DAG represents the conditional independences that can be read off by (*PMC*) and all the conditional independences that are entailed by them. Two DAGs are equivalent iff they represent the same set of conditional independences. It follows directly from the observational equivalence theorem (Pearl 2000, p. 19 with reference to Verma and Pearl 1990) that the DAG in figure 2 is equivalent to a DAG with the arrow from *aphids* to *fungus* reversed, but in all other respects the same.

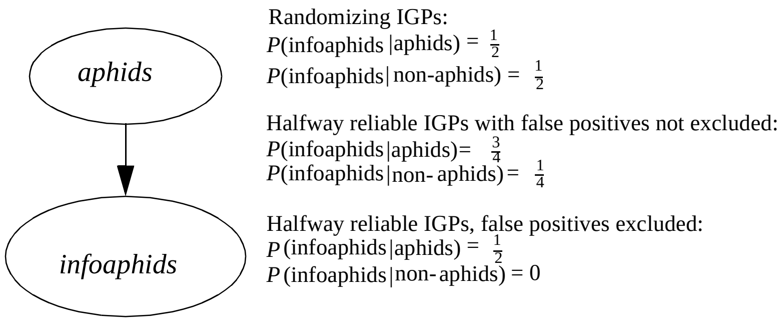


Figure 3: Information gathering processes (IGPs) that are believed not to be reliable

Similarly for the link from *fungus* to *infofungus*. We can now construct a 3 by 3 matrix of Bayesian networks with interpretations (i) through (iii) of "dubious information gathering processes" in its rows and with reclusive, indifferent, or sociable pest in its columns. As an illustration, we zoom in on the Bayesian network in the entry for reclusive pests and halfway reliable information gathering processes with false positives.

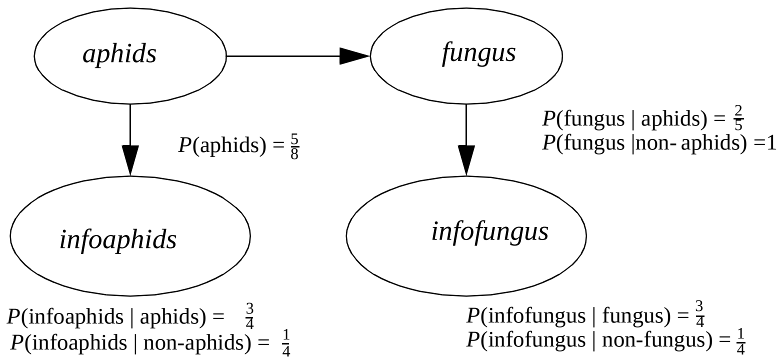


Figure 4: Reclusive pests and halfway reliable IGPs with false positives not excluded

What is so great about Bayesian networks? A network contains information about the independence relations between the variables, probability assignments for each root node, and conditional probability assignments for each child node. A foundational theorem in the theory of Bayesian networks states that a joint probability distribution over any combination of values of the variables in the network is equal to the product of the probabilities and conditional probabilities for these values as expressed in the network (Neapolitan 1990, 162–4). For example, suppose we are interested in the joint probability of aphids, non-fungus, infoaphid and infofungus for reclusive pests and halfway reliable information gathering processes. We can read the joint probability directly off of figure 4: *P*(aphids, non-fungus, infoaphids, infofungus) = *P*(aphids) *P*(non-fungus|aphids) *P*(infoaphids|aphids) *P*(infofungus|non-fungus) = $(\frac{5}{8})\ (\frac{3}{5})\ (\frac{3}{4})\ (\frac{1}{4})$. Standard probability calculus teaches us how to construct marginal distributions out of joint distributions and subsequently conditional distributions out of marginal distributions. But there is no need to get out your pocket calculator. Download your favourite Bayesian network software from the internet, pay the royalties and the results are at your fingertips. We are interested in the posterior joint probability that our tree has aphids and fungus, that is, the joint probability that our tree has aphids and fungus conditional on the fact that this information has gained access to our information set: *P**(aphids, fungus) = *P*(aphids, fungus|infoaphids, infofungus). So we enter infoaphids and infofungus as evidence in the network, let the evidence propagate through the network and read off the joint posterior probability of aphids and fungus: an exercise in computational epistemology.

The results are presented in table 1:

P^*(aphids, fungus)	*Reclusive*	*Indifferent*	*Sociable*
Randomizing IGPs	.25	.25	.25
Halfway reliable IGPs, false positives not excluded	.50	.5625	.6750
Halfway reliable IGPs, false positives excluded	1	1	1

Table 1: Posterior joint probabilities for IGPs believed to be not reliable

They are not too surprising. Under the background condition that we believe our information gathering processes to be no better than randomizing processes, the coherence of the information set is a red herring. If we could just as well have tossed a coin to determine whether our tree has aphids or whether our tree has fungus, then there is nothing to be learned from the information that we received: the joint probability of aphids and fungus remains unchanged, whether aphids and fungus are reclusive, indifferent, or sociable pests. Under the background condition that we believe our information gathering processes to be halfway reliable and false positives not to be excluded, the posterior joint probability of aphids and fungus is a positive function of the coherence of the information set: after we have received the information that our tree has aphids and fungus, it is indeed more probable that our tree actually has aphids and fungus when aphids and fungus are sociable pests than when they are indifferent pests and it is indeed more probable when they are indifferent pests than when they are reclusive pests. But under the background condition that we believe our information gathering processes to be halfway reliable but false positives to be excluded, the coherence of the information set is a red herring again: after we have received the information that our tree has aphids and fungus, we can rest assured that our tree has aphids and fungus, whether aphids and fungus are reclusive, indifferent or sociable pests. So far, the coherentist has scored one point and lost two points out of three. Of course, a set of Bayesian networks is not a proof. Theorem 1 in the appendix provides a proof of this result.

5. We Do Not Believe That Our Information Gathering Processes Are Reliable

So let us now turn to interpretations (iv) and (v) of "dubious information gathering processes". On both interpretations, we do not believe that our information gathering processes are reliable: there is some chance that they are (fully) reliable, and there is some chance that they are no better than randomizing processes. To model these interpretations we need to add at least two more nodes to our network. The variable *relinfoaphids* can take on two values, viz. relinfoaphids (i.e. the proposition that our information gathering process for aphids is reliable) and non-relinfoaphids (i.e. the proposition that our information gathering process for aphids is no better than a randomizing process). Whether we receive the information that our tree has aphids or not is directly causally influenced by whether our tree has aphids or not and by whether our information gathering process for aphids is reliable or not. Hence there is an arrow from *aphids* and an arrow from *relinfoaphids* entering into *infoaphids* with the following constraints on conditional probability assignments:

(15) $P(\text{infoaphids}|\text{relinfoaphids, aphids}) = 1;$

(16) $P(\text{infoaphids}|\text{relinfoaphids, non-aphids}) = 0;$

(17) $1 > P(\text{infoaphids}|\text{non-relinfoaphids, aphids}) = q$
$= P(\text{infoaphid}|\text{non-relinfoaphid, non-aphid}) > 0.$

To keep things simple, we let the information gathering processes for aphids and fungus be equally informative. Hence,

(18) P(infofungus | relinfofungus, fungus) = 1;

(19) P(infofungus | relinfofungus, non-fungus) = 0;

(20) $1 >$ P(infofungus | non-relinfofungus, fungus) = q
= P(infofungus|non-relinfofungus, non-fungus) > 0.

On interpretation (iv), we stipulate that there is some chance that the information gathering processes are reliable and some chance that they are not reliable. Furthermore, equal informativeness requires that the chance that they are reliable for aphids equals the chance that they are reliable for fungus:

(21) $1 >$ P(relinfoaphids) = p = P(relinfofungus) > 0.

We can now model interpretation (iv) in a Bayesian network. In figure 5 we have inserted some definite numbers: information gathering processes are just as likely to be fully reliable processes as to be no better than randomizing processes (p = .50), and a process that is no better than a randomizing process is like a coin flip (q = .50). The values are uniform across the scenarios of reclusive, indifferent, and sociable pests.

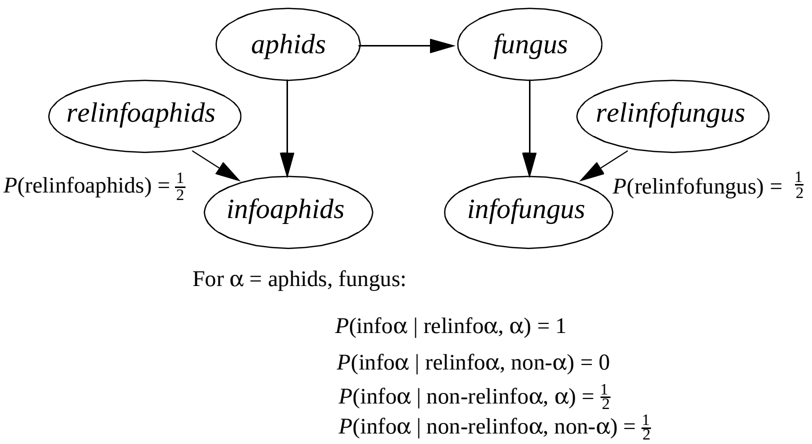


Figure 5: IGPs not believed to be reliable with independent quality

On interpretation (iv) the quality of one information gathering process is independent of the quality of the other information gathering process. We could conceive of this as follows. Suppose that there is a bag of aphid checking devices, half of them being fully reliable, half of them being randomizers. We pick one at random out of the bag. Same story for fungus checking devices. In this case our information gathering processes would be independent in the sense that the reliability of the fungus checking device is irrelevant to the reliability of the aphid checking advice:

(22) $\quad$ *relinfoaphids* $\perp\!\!\!\perp$ *relinfofungus*.

Let us check whether this condition is indeed fulfilled in the Bayesian network. The standard way of doing so is to appeal to the d-separation criterion. (See our example in the preliminary to the appendix.) But also (*PMC*) establishes this result in a slightly roundabout manner. A root node in the network can be interpreted as a child node to the parent node which represents the propositional variable **T**, that is, the tautology. Then we can read off of the network that the child node *relinfoaphids* is indeed independent of its non-descendant *relinfofungus*, conditional on its parent node **T**, that is, *relinfoaphids* is independent of *relinfofungus tout court*.

Let us now turn to interpretation (v). On this interpretation, the quality of one information gathering process is maximally dependent on the quality of the other information gathering processes. We could conceive of this as follows. Suppose that there is a bag of information gatherers, some of them reliable, some of them non-reliable. Reliable information gatherers use only reliable information gathering processes, both for aphids and for fungus. Non-reliable information gatherers only use processes that are no better than randomizing processes, both for aphids and for fungus. We were picked out of the bag at random, but we do not know which group of information gatherers we belong to.

It is straightforward how to model this into a Bayesian network. Add one node to the network representing the variable *relinfogath*. This variable can take two values, viz. relinfogath (i.e. the proposition that we are reliable information gatherers) and non-relinfogath (i.e. the proposition that we are a non-reliable information gatherers). Whether our information gathering process for aphids is reliable or not is directly causally influenced by whether we are reliable information gatherers or not. Similarly for fungus. Hence, we draw an arrow from *relinfogath* to *relinfoaphids* and an arrow from *relinfogath* to *relinfofungus*. We assign a probability to the node *relinfogath* and conditional probabilities to the node *relinfoaphids*:

(23) $\quad P(\text{relinfogath}) = \text{s};$

(24) $\quad P(\text{relinfoaphids}|\text{relinfogath}) = 1;$

(25) $\quad P(\text{relinfoaphids}|\text{not-relinfogath}) = 0.$

Similarly, equal informativeness requires that the following conditional probabilities hold for the node *relinfofungus*:

(26) P(relinfofungus|relinfogath) = 1;

(27) P(relinfofungus|not-relinfogath) = 0.

In figure 6 we have set s at .50: we are just as likely to be reliable as non-reliable information gatherers. The values are uniform across the scenarios of reclusive, indifferent, and sociable pests.

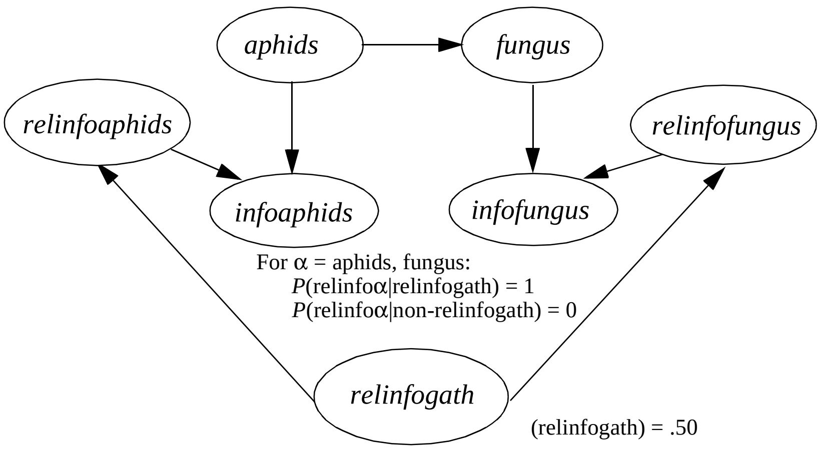


Figure 6: IGPs not believed to be reliable with maximally dependent quality

The independence in (22) is no longer implied by this new Bayesian network. We have added a joint parent, viz. *relinfogath*, to *relinfoaphids* and *relinfofungus* in the new network: hence (*PMC*) supports the independence of *relinfoaphids* on its non-descendant node *relinfofungus*, conditional on its parent *relinfogath*, but it no longer establishes the independence of *relinfoaphid* and *relinfofungus tout court*. (For a more rigorous proof on the basis of the d-separation criterion, see our example in the preliminary to the appendix.)

We can construct a 2 by 3 matrix of Bayesian networks for interpretations (iv) and (v) of "dubious information gathering processes" in its rows and reclusive, indifferent, and sociable pests in its columns. Once again, we enter infoaphids and infofungus as evidence in the network, let the evidence propagate throughout the network, and read off the posterior joint probabilities for aphids and fungus. What is there to say? Table 2 summarizes the results:

P^*(aphids, fungus)	*Reclusive*	*Indifferent*	*Sociable*
Independent quality	.50	.5625	.675
Maximally dependent quality	.6250	.6250	.6250

Table 2: Posterior joint probabilities for IGPs not believed to be reliable

It is a tie for the coherentist. When we do not believe that our information gathering processes are reliable, but the quality of the processes is independent, then the posterior joint probability of aphids and fungus is a positive function of the coherence of the information set. When we do not believe that our information gathering processes are reliable, but the quality of the processes is maximally dependent, then the coherence of the information set is once again a red herring. The posterior joint probability of aphids and fungus rises substantially, but the coherence of the information set is immaterial to the size of this gain.

Bayesian networks are not proofs. Theorem 2 in the appendix proves that the posterior joint probability is a positive function of the coherence of the information set on interpretation (iv) of “dubious information gathering processes”. Theorem 3 proves that the posterior joint probability remains unaffected by the coherence of the information set on interpretation (v) of “dubious information gathering processes”.

How are we to understand these results? We are dealing twice with sceptics about the quality of our information gathering processes. On interpretation (iv), our sceptic is also an extreme sceptic about the power of induction, while on interpretation (v), our sceptic is extremely gullible when it comes to the power of induction. To add some drama to the picture, let us divide our sceptics in two schools and add some gratuitous cosmological spice.

A sceptic of the anti-inductivist school makes the following claim: we are in doubt about the reliability of our information gathering processes; in addition, the supposition that all but one of our information gathering processes are reliable would not affect the chance that we assign to that very last one being reliable as well. His credo is that the creator is a gambler: before creating each one of our senses, the creator flipped a coin to determine whether to make it reliable or not. Sceptics from this school should be impressed by coherence: the more coherent our information set turns out to be, the more credible its content after the reports have come in. A sceptic of the inductivist school makes the following claim: we are in doubt about our information gathering processes; but the supposition that one is reliable would make us conclude that the other ones are reliable as well. Her credo is that the creator is principled: either he made all of our senses reliable, or he made all of our senses unreliable. Sceptics from this school can turn their back on coherence: the coherence of the information set is irrelevant to the credibility of the content of the information set after the reports have come in.

If you are like us, you do not want to join either one of these schools. With the person on the street, we believe that if we come to learn that some of our information gathering processes are reliable, then it makes us feel somewhat more confident that our other information gathering processes are reliable as well: it is not entirely irrelevant to our belief that our other senses are reliable as well, as the sceptics of the anti-inductivist school preach, nor is it sufficient to convince us that our other senses are reliable as well, as the sceptics of the inductivist school preach. Common sense teaches that our information gathering processes are moderately dependent. So what about coherence? Is the coherence of the information set relevant to those who abide by common sense? There is a simple way of modelling the common sense view in a Bayesian network. We will take the gratuitous cosmology of the schools as our inspiration.

The Bayesian network in figure 5 is the logo of the anti-inductivist school. The Bayesian network in figure 6 is the logo of the inductivist school. We can express precisely the same information that is present in the Bayesian network in figure 5 by means of a Bayesian network that has the same structure as the network in figure 6, if we adjust the conditional probabilities of *relinfoaphids* and of *relinfofungus* on *relinfogath*. Remember our gratuitous cosmology: let there be a creator who flipped a coin for each of our senses to determine whether it will be reliable or not; we add that the creator is either a god or a demon, but gods and demons are all of a kind when it comes to determining the reliability of our senses. As a mnemonic aid, substitute *creator* for *relinfogath* and let this variable take on two values, viz. god and demon. We insert the values for the anti-inductivist school into figure 7: $P(\text{god}) = P(\text{demon}) = .50$ and $P(\text{relinfoaphids}|\text{god}) = P(\text{relinfoaphids}|\text{demon}) = .50$ and similarly for relinfofungus. It now represents the same substantial information as figure 5: the quality of our information gathering processes is independent.

Similarly, we can add some gratuitous cosmological spice to the inductivist school. We do not know whether our creator is a god or a demon. Let $P(\text{god}) = P(\text{demon}) = .50$. If our creator is a god, then all of our senses are reliable. If our creator is a demon, then all of our senses are unreliable: $P(\text{relinfoaphids}|\text{god}) = P(\text{relinfofungus}|\text{god}) = 1$ and $P(\text{relinfoaphids}|\text{demon}) = P(\text{relinfofungus}|\text{demon}) = 0$. If we insert these values into figure 7, then it represents the same substantial information as figure 6: the quality of our information gathering processes is maximally dependent.

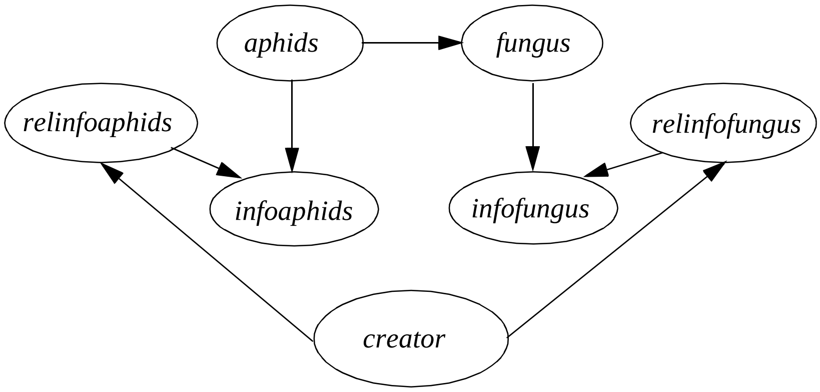


Figure 7: IGPs not believed to be reliable with gratuitous cosmology

Now it is easy to see how to model the common sense view. The person on the street also believes that the creator may have been a god or may have been a demon. Let *P*(god) = *P*(demon) = .50. Gods are more benevolent than demons, but they are not all-benevolent: for each sense organ there is a three in four chance that it is reliable: *P*(relinfoaphids|god) = *P*(relinfofungus|god) = .75. Demons are more malevolent than gods, but they are not all-malevolent: for each sense organ there is a one in four chance that it is reliable: *P*(relinfoaphids|demon) = *P*(relinfofungus|demon) = .25. If we insert these values into figure 7, it represents the common sense view that the quality of our information gathering processes is moderately dependent: coming to learn that some of our senses are reliable gives us good reason to believe that the creator was a god rather than a demon, and hence we can be somewhat more confident that our other senses will be reliable as well.

So let us run the networks again. With the values of the anti-inductivist schools inserted, the network in figure 7 yields precisely the same results as the network in figure 5: coherence matters to the posterior joint probability when the quality of the information gathering processes is independent, whether supernatural beings are in play or not. With the values of the inductivist school inserted, the network in figure 7 yields precisely the same results as the network in figure 6: coherence does not matter a bit to the posterior joint probability when the quality of the information gathering processes is maximally dependent, whether we baptize the bottom node *relinfogath* or *creator*. With the values for the common sense view inserted, the network in figure 7 yields the following result: when the quality of our information gathering processes is only moderately dependent, coherence does not assert its full force, but it still does raise the posterior joint probability.

P^*(aphids, fungus)	*Reclusive*	*Indifferent*	*Sociable*
Independent quality	.50	.5625	.675
Moderately dependent quality	.5572	.5781	.6606
Maximally dependent quality	.625	.625	.625

Table 3: Posterior joint probabilities for IGPs not believed to be reliable, including the common sense view of dependent quality

It is not too hard to understand why coherence is a red herring in the inductivist school. Suppose that we believe that our information gathering processes are fully reliable. Then coherence clearly has no impact: if we receive fully reliable reports that our tree has aphids and that our tree has fungus, then we can be certain that our tree has aphids and fungus, whether aphids and fungus are reclusive, indifferent or sociable pests.[5] What is believed in the inductivist school is that either all our information gathering processes are no better than randomizing processes or all our information gathering processes are fully reliable. If the former is the case, then we know from interpretation (i) that coherence plays no role. If the latter is the case, then we also know that coherence plays no role. Hence, it should be no surprise that on interpretation (v) coherence is a red herring.

One might object that if we really do not believe that our information gathering processes are reliable, then we should allow some chance that they are fully reliable, some chance that they are no better than randomizing processes, but also some chance that they are halfway reliable and yield false positives. This will make a difference to the role of coherence in the inductivist school. Remember that interpretation (ii) did leave room for false positives and that coherence did make a difference to the posterior joint probability. It is to be expected that as soon as the inductivist allows that all our information gathering processes are such that they may yield false positives, coherence does start making a difference to the pos-

[5] It is easy to prove this. Add the following line to theorem 1 in the appendix: (iv) If $\boldsymbol{P}(\text{ia}|\text{a}) = 1$ and $\boldsymbol{P}(\text{ia}|\bar{\text{a}}) = 0$, then r = 1. This result follows directly from line (6) and (8) of the theorem.

terior joint probability. It is easy to model this. Suppose that our creator is either a god, an angel or a demon. If it is a god, our information gathering processes are fully reliable; if it is an angel, they are halfway reliable and occasionally yield false positives; if it is a demon, they are no better than randomizing processes. The variable *creator* can take on three values, viz. god, angel and demon. The variable *relinfoaphids* can take on three values, viz. reliableaphids, randomaphids and halfwayaphids. Similarly for *relinfofungus*. We reconstruct the Bayesian network in figure 7 and insert some definite values. For α = aphids, fungus

$P(\text{god}) = \frac{1}{3}$	$P(\text{reliable}\alpha \mid \text{god}) = 1$	$P(\text{info}\alpha \mid \text{reliable}\alpha, \alpha) = 1$
$P(\text{angel}) = \frac{1}{3}$	$P(\text{halfway}\alpha \mid \text{angel}) = 1$	$P(\text{info}\alpha \mid \text{reliable}\alpha, \text{non-}\alpha) = 0$
$P(\text{demon}) = \frac{1}{3}$	$P(\text{random}\alpha \mid \text{demon}) = 1$	$P(\text{info}\alpha \mid \text{halfway}\alpha, \alpha) = \frac{3}{4}$
		$P(\text{info}\alpha \mid \text{halfway}\alpha, \text{non-}\alpha) = \frac{1}{4}$
		$P(\text{info}\alpha \mid \text{random}\alpha, \alpha) = \frac{1}{2}$
		$P(\text{info}\alpha \mid \text{random}\alpha, \text{non-}\alpha) = \frac{1}{2}$

Table 4: IGPs not believed to be reliable admitting the possibility of halfway reliability with false positives.

Running the network yields the following posterior joint probabilities:

	Reclusive	*Indifferent*	*Sociable*
P^*(aphids, fungus)	.58	.60	.64

Table 5: Posterior joint probabilities for IGPs not believed to be reliable admitting the possibility of halfway reliability with false positives

These results confirm our expectations: if one admits the possibility that our information gathering processes are halfway reliable and yield false positives, then coherence does matter even for the sceptic of the inductivist school.

7. Discussion

To introduce the role of dubious information gathering processes in coherentism we started with some quotes from C.I. Lewis and Bonjour. Let us revisit these quotes to show how much confusion there is about this particular ingredient of the coherentist canon. Lewis himself is a weak foundationalist about memory: our memory beliefs must have at least some independent credibility which can then be amplified on coherence grounds. Bonjour defends his theory against the charge that his version of coherentism is also a form of weak foundationalism. To distance himself from Lewis, he claims that the coherence of a set of propositions will render them credible even under conditions that suggest extremely dubious information gathering processes. A description of such conditions can be lifted from the Bonjour passage that we quoted in section 1:

(a) "… no antecedent degree of warrant or credibility is required";

(b) "… individual reports initially have a high degree of negative credibility, that is are much more likely to be false than true…";

(c) "… all of the witnesses are known to be habitual liars".

Claim (b) is clearly true. In our terminology, it says that the posterior probability of each proposition in the information set after a corresponding report was received may still be lower than .50. It is easy to construct an example in which coherence is effective so that the respective posterior probabilities of every single proposition after the corresponding report is in are lower than .50, while the posterior joint probability of the propositions after *all* the reports are in is close to 1.

Claim (c) is difficult to assess because it is not clear what an habitual liar is. But there is a charitable reading on which coherence is effective even when the information is obtained from habitual liars. Suppose that an habitual liar is someone who more often than not provides false information about items within his area of expertise. We adopt our interpretation (ii) of "dubious information gathering processes", that is, the interpretation on which the information gathering processes are halfway reliable and yield false positives. Suppose that $P(\text{infoaphids}|\text{aphids}) = .20 > (\text{infoaphids}|\text{non-aphids}) = .05$ and that the aphid specialist's only alternative to reporting that the tree has aphids is that the tree has malaria—a report which is bound to be false. Then the aphids specialists lies 80% of the time when the tree has aphids and 100% of the time when the tree does not have aphids. Similarly for fungus. The aphids and fungus specialists are habitual liars, but remember that coherence is effective on interpretation (ii) of "dubious information gathering processes".

Claim (a) is also quite lean and open to interpretation. What it cannot mean is that the prior joint probability of the propositions in the information set is set at 0, since no amount of information, whether coherent or not, could change such an assignment. Then what is Bonjour up to? Here is a charitable reading. Bonjour may be making a claim about the quality of our information gathering processes: for coherence to be effective, it is not required that we believe our information gathering to be reliable. This reading corresponds to interpretations (iv) and (v) of "dubious information gathering processes". As we have seen, the coherentist canon does hold up as long as there is a mix of doxastically possible worlds in which the information gathering processes are reliable and worlds in which they are no better than randomizing processes and as long as the quality of the information gathering processes is not maximally dependent: we are in interpretation (iv) or on the continuum between interpretations (iv) and (v) of "dubious information gathering processes". Hence, claim (a) can be made true on this reading.

But this reading may be too charitable. Remember that Bonjour is presenting an argument against C.I. Lewis's claim that "[for] any one of [the] reports taken singly, the extent to which it confirms what is reported may be slight" (Lewis 1946, p. 346). Lewis seems to imply that coherence is effective only if the posterior probability of a single proposition after the corresponding report (and only the corresponding report) has come in exceeds its prior probability ever so slightly. So if Bonjour is to provide a counter to Lewis, then what he must be claiming is that coherence is effective also when the posterior probability of a single proposition after the corresponding report has come in equals the prior probability. But this condition is tantamount to the condition that our information gathering processes are such that the probability of a report given that what is reported is true is equal to the probability of a report given that what is reported is false (under innocent assumptions).[6] This is precisely interpretation (i), that is, the interpretation under which we believe our information gathering processes to be no better than randomizing processes. As we have seen, coherence is a red herring under this interpretation. Hence, on this reading, claim (a) is false.

Michael Huemer (1997) presents a Bayesian argument against Bonjour that is similar to our argument in the last paragraph and he starts tolling the bells on coherentism. We have had no interest in tolling any bells. We laid out a rough coherentist canon and determined under what refinements the canon is true and under what refinements the canon is

[6] Let α be the content of the report and infoα be the claim that there is a report to the effect that α. Then, by the probability calculus, $P(\alpha \mid \text{info}\alpha) = P(\alpha)$ iff $P(\text{info}\alpha \mid \alpha) = P(\text{info}\alpha)$ iff $P(\text{info}\alpha \mid \alpha) = P(\text{info}\alpha \mid \text{non-}\alpha)$, assuming that $P(\alpha)>0$ and $P(\text{info}\alpha)>0$.

false. If someone wishes to argue that *pure* coherentism should really embrace one or another refinement, then let him be our guest. If camps must be picked, then the summary of our argument in the next paragraph should be your guide: believers can stake out the true faith; reformers can load some untenable interpretation on coherentist shoulders. We do not wish to be cynical: if good arguments are forthcoming for or against some interpretation, we are willing to convert. But as of now, we are happy sitting on the fence.

The coherentist canon requires a much finer specification of what is meant by "dubious information gathering processes". If it means that we believe our information gathering processes to be no better than randomizing processes or halfway reliable processes that do not yield false positives, the canon is false. If it means that we believe that our information gathering processes are halfway reliable and yield false positives, the canon is true. If it means that we do not believe that our information gathering processes are reliable, then the question is what other options we leave open. If we leave it open that our information gathering processes are halfway reliable and yield false positives, then the canon is true. But if we do not leave this option open and only leave it open that our information gathering processes are fully reliable or are no better than randomizing processes, then a new question is in order: do we take the reliability of some information gathering process to be evidence for the reliability of others? If we do so to the greatest extent, then the canon is false. If we do so to a lesser extent or not at all, then the canon is true. So pick your pile of wood and fire up the stakes![7]

Appendix

Preliminary: The Parental Markov Condition (*PMC*) (see § 3) permits us to read a number of conditional independencies from the network. But these are usually not the only conditional independencies in the network. Additional conditional independences can be derived by means of the semi-graphoid axioms. These axioms were developed by Dawid (1979) and Spohn (1980) and are presented in Pearl (1988, pp. 82–90). Alternatively, we can appeal to the d-separation criterion to read off the same

[7] Thanks to George Bealer, Stephan Hartmann, Michael Huemer, Michelle Montague and the referees of this journal for helpful comments and suggestions. Luc Bovens's research was supported by an Alexander von Humboldt Foundation Fellowship and by a grant of the National Science Foundation, Science and Technology Studies(SES 00-80580). Erik J. Olsson's research was supported by the DFG (*Deutsche Forschungsgemeinschaft*) as a contribution to the project *Logik in der Philosophie*.

additional conditional independences (Neapolitan 1990, 202–7 with reference to Verma and Pearl 1988). Let α, β, γ be sets of variables in the network. Then the following condition is a sufficient condition to determine that $\alpha \perp\!\!\!\perp \beta \mid \gamma$. Construct a matrix of the variables in α by the variables in β. Focus on some entry (*A*, *B*) with $A \in \alpha$ and $B \in \beta$ in this matrix. Trace all the paths between the variables *A* and *B*. A path is blocked just in case it contains at least one serial node or divergent node which is in γ or at least one convergent node which is such that neither it nor any of its descendants are in γ. (A serial node is a node which has an incoming and an outgoing arrow; a divergent node is a node which has two outgoing arrows; a convergent node is a node which has two incoming arrows.) Check whether all the paths between *A* and *B* are blocked. If so, then move on to the next entry in the matrix. If for each entry in the matrix all the paths are blocked, then the conditional independence $\alpha \perp\!\!\!\perp \beta \mid \gamma$ holds.

Example: We check whether *relinfoaphids* $\perp\!\!\!\perp$ *relinfofungus* (i.e. {*relinfoaphids*} $\perp\!\!\!\perp$ {*relinfoaphids*} | $\varnothing$) holds in the network in figure 5. There is a single path between *relinfoaphids* and *relinfofungus*, viz. over *infoaphids*-*aphids*-*fungus*-*infofungus*. Since the convergent node *infoaphids* $\notin \varnothing$ (or, since the convergent node *infofungus* $\notin \varnothing$), this path is blocked and we may conclude that *relinfoaphids* $\perp\!\!\!\perp$ *relinfofungus*. Does this independence hold in figure 6? There are two paths between *relinfoaphids* and *relinfofungus*, viz. over *infoaphids*-*aphids*-*fungus*-*infofungus* and over *relinfogath*. We focus on the latter path. Since *relinfogath* is a divergent node and *relinfogath* $\notin \varnothing$, this path is not blocked and we can no longer conclude that *relinfoaphids* $\perp\!\!\!\perp$ *relinfofungus*.

Theorems: We abbreviate the variables *aphids* and *fungus* as *a* and *b* (respectively); *infoaphids* and *infofungus* as *ia* and *ib*; *relinfoaphids* and *relinfofungus* as *ria* and *rib*; and *relinfogath* as *rig*. Each variable can take on two values: the variable *a* can take on the values a and $\bar{\text{a}}$ etc. Let ***P*** range over the probability distributions *P* and *P*'. We write $P \succ P'$ iff {a,b} is more coherent on *P* than on *P*', as defined in section 1. We are interested in the ratio r = $P(\text{a,b}|\text{ia,ib})/P'(\text{a,b}|\text{ia,ib})$.

Theorem 1: Assuming (A1) the independencies embedded in the network in figure 2; (A2) $0 < P(\text{a,b}) = P'(\text{a,b}) < 1$ (*fixed joint priors*); (A3) $\boldsymbol{P}(\text{ia}|\text{a}) = \boldsymbol{P}(\text{ib}|\text{b})$; $\boldsymbol{P}(\text{ia}|\bar{\text{a}}) = \boldsymbol{P}(\text{ib}|\bar{\text{b}})$ (*equally informative IGPs*); and (A4) $P(\text{ia}|a) = P'(\text{ia}|a)$ (*uniform IGPs across distributions*), the following holds: $P \succ P'$ entails that

(i) If $\boldsymbol{P}(\text{ia}|\text{a}) = \boldsymbol{P}(\text{ia}|\bar{\text{a}}) > 0$, then r = 1;
(ii) If $\boldsymbol{P}(\text{ia}|\text{a}) > \boldsymbol{P}(\text{ia}|\bar{\text{a}}) > 0$, then r>1;
(iii) If $1 > \boldsymbol{P}(\text{ia}|\text{a}) > \boldsymbol{P}(\text{ia}|\bar{\text{a}}) = 0$, then r = 1.

Proof: A1 permits us to read off the following independences by means of the d-separation criterion:

(1) $ia \perp\!\!\!\perp ib \mid a,b$; $ia \perp\!\!\!\perp b \mid a$; $ib \perp\!\!\!\perp a \mid b$.

Apply Bayes theorem to r, work out the independences as specified in (1) and the identities as specified in A3 and A4. Set $\alpha = P(\text{ia}|\text{a})P(\text{ib}|\text{b})$; $\beta = P(\text{ia}|\bar{\text{a}})P(\text{ib}|\text{b})$; $\gamma = P(\text{ia}|\text{a})P(\text{ib}|\bar{\text{b}})$; $\delta = P(\text{ia}|\bar{\text{a}})P(\text{ib}|\bar{\text{b}})$; $w = P(\text{a, b})$ and $w' = P'(\text{a, b})$; $x = P(\bar{\text{a}}, \text{b})$ and $x' = P'(\bar{\text{a}}, \text{b})$; $y = P(\text{a}, \bar{\text{b}})$; $y' = P'(\text{a}, \bar{\text{b}})$; $z = P(\bar{\text{a}}, \bar{\text{b}})$ and $z' = P'(\bar{\text{a}}, \bar{\text{b}})$. This yields:

(2) $$r > 1 \Leftrightarrow \frac{\alpha w/(\alpha w + \beta x + \gamma y + \delta z)}{\alpha w'/(\alpha w' + \beta x' + \gamma y' + \delta z')} > 1$$

By A2, w = w'. Since αw, $\alpha w + \beta x + \gamma y + \delta z$, $\alpha w + \beta x' + \gamma y' + \delta z' > 0$,

(3) $r > 1 \Leftrightarrow \beta x + \gamma y + \delta z < \beta x' + \gamma y' + \delta z$.

By the probability calculus, $z = 1 - (w+x+y)$ and $z' = 1 - (w+x'+y')$ and again, by A2, w = w'. Substitution and some algebraic manipulation yields:

(4) $r > 1 \Leftrightarrow (\beta - \delta)(x - x') < (\gamma - \delta)(y' - y)$.

Returning to longhand notation:

(5) $r > 1 \Leftrightarrow$

$$P(\text{ia}|\bar{\text{a}})\,[P(\text{ib}|\text{b}) - P(\text{ib}|\bar{\text{b}})][P(\bar{\text{a}}, \text{b}) - P'(\bar{\text{a}},\text{b})] < P(\text{ib}|\bar{\text{b}})\,[P(\text{ia}|\text{a}) - P(\text{ia}|\bar{\text{a}})]\,[P'(\text{a}, \bar{\text{b}}) - P(\text{a},\bar{\text{b}})].$$

Similarly,

(6) $r = 1 \Leftrightarrow$

$$P(\text{ia}|\bar{\text{a}})\,[P(\text{ib}|\text{b}) - P(\text{ib}|\bar{\text{b}})]\,[P(\bar{\text{a}},\text{b}) - P'(\bar{\text{a}},\text{b})] = P(\text{ib}|\bar{\text{b}})\,[P(\text{ia}|\text{a}) - P(\text{ia}|\bar{\text{a}})]\,[P'(\text{a}, \bar{\text{b}}) - P(\text{a}, \bar{\text{b}})].$$

Now assume $P > P'$. By A2,

(7) $P(\text{a} \vee \text{b}) < P'(\text{a} \vee \text{b})$.

Again by A2,

(8) $P(\bar{\text{a}},\text{b}) - P'(\bar{\text{a}},\text{b}) < P'(\text{a},\bar{\text{b}}) - P'(\text{a},\bar{\text{b}})$.

By A3, part (ii) of the theorem follows from (5) and (8) and parts (i) and (iii) follow from (6) and (8).

Theorem 2: Assuming (B1) the independences embedded in the network in figure 5; (B2) $P(\text{a,b}) = P'(\text{a,b})$ (*fixed joint priors*); (B3) $\boldsymbol{P}(\text{ia}|\text{a},\text{ria}) = \boldsymbol{P}(\text{ib}|\text{b},\text{rib})$; $\boldsymbol{P}(\text{ia}|\bar{\text{a}},\text{ria}) = \boldsymbol{P}(\text{i,b}|\bar{\text{b}},\text{rib})$; $\boldsymbol{P}(\text{ia}|\text{a},\overline{\text{ria}}) = \boldsymbol{P}(\text{ib}|\text{b},\overline{\text{rib}})$; $\boldsymbol{P}(\text{ia}|\bar{\text{a}},\overline{\text{ria}}) = \boldsymbol{P}(\text{ia}|\bar{\text{b}},\overline{\text{rib}})$; $\boldsymbol{P}(\text{ria}) = \boldsymbol{P}(\text{rib})$ (*equally informative IGPs*); (B4) $P(\text{ia}|a, ria) = P'(\text{ia}|a, ria)$; $P(ria) = P'(ria)$ (*uniform IGPs across* *distributions*), the following holds: $P > P'$ entails that, if

(i) $1 > \mathbf{P}(\text{ia}|\text{a},\overline{\text{ria}}) = \mathbf{P}(\text{ia}|\bar{\text{a}},\overline{\text{ria}}) > 0$

(ii) $\boldsymbol{P}(\text{ia}|\text{a},\text{ria}) = 1$ and $\boldsymbol{P}(\text{ia}|\bar{\text{a}},\text{ria}) = 0$ and

(iii) $1 > \boldsymbol{P}(\text{ria}) > 0$,

then $r > 1$.

Proof: First we will prove that the assumptions in theorem 2, that is, B1 through B4, entail the assumptions in theorem 1, that is, A1 through A4. First, consider the assumptions A1 and B1. Since the addition of the variable *ria* and *rib* to figure 2 in figure 5 leaves all the old paths intact and creates no new paths between any of the variables in figure 2, all the independences that can be read off from figure 2 by means of the d-separation criterion can also be read off from figure 5. Second, A2 is identical to B2. Third, we will prove that B1 and B3 jointly entail A3. The following are two facts of the probability calculus: for all values of *a* and *b*,

(1) $\boldsymbol{P}(\text{ia}|a) = \boldsymbol{P}(\text{ia}|a,\text{ria})\boldsymbol{P}(\text{ria}|a) + \boldsymbol{P}(\text{ia}|a,\overline{\text{ria}})\,\boldsymbol{P}(\overline{\text{ria}}|a)$;

(2) $\boldsymbol{P}(\text{ib}|b) = \boldsymbol{P}(\text{ib}|b,\text{rib})\boldsymbol{P}(\text{rib}|b) + \boldsymbol{P}(\text{ib}|b,\overline{\text{rib}})\boldsymbol{P}(\overline{\text{rib}}|b)$.

B1 permits us to read off the following independences by means of the d-separation criterion:

(3) $ria \perp\!\!\!\perp a$; $rib \perp\!\!\!\perp a$.

In (1) and (2) we work out the independences as specified in (3):

(4) $\boldsymbol{P}(\text{ia}|a) = \boldsymbol{P}(\text{ia}|a,\text{ria})\boldsymbol{P}(\text{ria}) + \boldsymbol{P}(\text{ia}|\text{a},\overline{\text{ria}})\boldsymbol{P}(\overline{\text{ria}})$;

(5) $\boldsymbol{P}(\text{ib}|b) = \boldsymbol{P}(\text{ib}|b,\text{rib})\boldsymbol{P}(\text{rib}) + \boldsymbol{P}(\text{ib}|b,\overline{\text{rib}})\boldsymbol{P}(\overline{\text{rib}})$.

By (B3) it follows from (4) and (5):

(6) $\boldsymbol{P}(\text{ia}|\text{a}) = \boldsymbol{P}(\text{ib}|\text{b})$ and $\boldsymbol{P}(\text{ia}|\bar{\text{a}}) = \boldsymbol{P}(\text{ib}|\bar{\text{b}})$.

Fourth, by a parallel argument, it can be shown that B1 and B4 jointly entail A4.

Since the assumptions in theorem 2 entail the assumptions in theorem 1, we may borrow the following result from theorem 1:

(7) $P > P'$ entails that, if $\boldsymbol{P}(\text{ia}|\text{a}) > \boldsymbol{P}(\text{ia}|\bar{\text{a}}) > 0$, then $r > 1$.

To complete the proof, we only need to show that (i), (ii) and (iii) in theorem 2 jointly entail that $P(\text{ia}|\text{a}) > P(\text{ia}|\bar{\text{a}}) > 0$. From (ii) and (iii), it follows that:

(8) $\boldsymbol{P}(\text{ia}|\text{a},\text{ria})\boldsymbol{P}(\text{ria}) > \boldsymbol{P}(\text{ia}|\bar{\text{a}},\text{ria})\,\boldsymbol{P}(\text{ria}) = 0$.

From (i), (iii) and (8), it follows that:

(9) $\boldsymbol{P}(\text{ia}|\text{a},\text{ria})\boldsymbol{P}(\text{ria}) + \boldsymbol{P}(\text{ia}|\text{a},\overline{\text{ria}})\,\boldsymbol{P}(\overline{\text{ria}}) > \boldsymbol{P}(\text{ia}|\bar{\text{a}},\text{ria})\,\boldsymbol{P}(\text{ria})$
$+ \boldsymbol{P}(\text{ia}|\bar{\text{a}},\overline{\text{ria}})\,\boldsymbol{P}(\overline{\text{ria}}) > 0$.

By the independence in (3):

(10) $\boldsymbol{P}(\text{ia}|\text{a},\text{ria})\boldsymbol{P}(\text{ria}|\text{a}) + \boldsymbol{P}(\text{ia}|\text{a},\overline{\text{ria}})\,\boldsymbol{P}(\overline{\text{ria}}|\text{a}) >$

$\boldsymbol{P}(\text{ia}|\bar{\text{a}},\text{ria})\boldsymbol{P}(\text{ria}|\bar{\text{a}}) + \boldsymbol{P}(\text{ia}|\bar{\text{a}},\overline{\text{ria}})\,\boldsymbol{P}(\overline{\text{ria}}|\bar{\text{a}}) > 0.$

By the probability calculus, (10) is equivalent to $\boldsymbol{P}(\text{ia}|\text{a}) > \boldsymbol{P}(\text{ia}|\bar{\text{a}}) > 0$, which concludes our proof.

Theorem 3: Assuming (C1) the independences embedded in the network in figure 6; (C2) $P(\text{a,b}) = P'(\text{a,b})$ (*fixed joint priors*); (C3) $\boldsymbol{P}(\text{ia}|\text{a},\text{ria}) = \boldsymbol{P}(\text{ib}|\text{b},\text{rib})$; $\boldsymbol{P}(\text{ia}|\bar{\text{a}},\text{ria}) = \boldsymbol{P}(\text{ib}|\bar{\text{b}},\text{rib})$; $\boldsymbol{P}(\text{ia}|\text{a},\overline{\text{ria}}) = \boldsymbol{P}(\text{ib}|\text{b},\overline{\text{rib}})$; $\boldsymbol{P}(\text{ia}|\bar{\text{a}},\overline{\text{ria}}) = \boldsymbol{P}(\text{ib}|\bar{\text{b}},\overline{\text{rib}})$***;*** $\boldsymbol{P}(\text{ria}|\text{rig}) = \boldsymbol{P}(\text{rib}|\text{rig})$ and $\boldsymbol{P}(\text{ria}|\overline{\text{rig}}) = \boldsymbol{P}(\text{rib}|\overline{\text{rig}})$ (*equally informative IGPs*) and (C4) $P(\text{ia}|a, ria) = P'(\text{ia}|a, ria)$; $P(\text{ria}|rig) = P'(\text{ria}|rig)$ and $P(\text{rig}) = P'(\text{rig})$ (*uniform IGPs across distributions*), the following holds: $P > P'$ entails that, if

(i) $1 > \boldsymbol{P}(\text{ia}|\text{a},\overline{\text{ria}}) = \boldsymbol{P}(\text{ia}|\bar{\text{a}},\overline{\text{ria}}) > 0$,

(ii) $\boldsymbol{P}(\text{ia}|\text{a},\text{ria}) = 1$ and $\boldsymbol{P}(\text{ia}|\bar{\text{a}},\text{ria}) = 0$, and

(iii) $\boldsymbol{P}(\text{ria}|\text{rig}) = 1$; $\boldsymbol{P}(\text{ria}|\overline{\text{rig}}) = 0$,

then r = 1.

Proof: By the probability calculus, for all values of *a* and *b*:

(1) $\boldsymbol{P}(\text{ia,ib}|a,b) = \boldsymbol{P}(\text{ia,ib}|a,b,\text{rig})\boldsymbol{P}(\text{rig}|a,b) +$
$\boldsymbol{P}(\text{ia,ib}|a,b,\overline{\text{rig}})\,\boldsymbol{P}(\overline{\text{rig}}|a,b).$

C1 permits us to read off the following independences by the d-separation criterion:

(2) $rig \perp\!\!\!\perp a,b$; $ia \perp\!\!\!\perp b|a,rig$; $ib \perp\!\!\!\perp a|b,rig$; $ia \perp\!\!\!\perp ib|a,b,rig$.

In (1) we work out the independences as specified in (2):

(3) $\boldsymbol{P}(\text{ia,ib}|a,b) = \boldsymbol{P}(\text{ia}|a,\text{rig})\boldsymbol{P}(\text{ib}|b,\text{rig})\boldsymbol{P}(\text{rig}) +$
$\boldsymbol{P}(\text{ia}|a,\overline{\text{rig}})\boldsymbol{P}(\text{ib}|b,\overline{\text{rig}})\boldsymbol{P}(\overline{\text{rig}}).$

We instantiate for a and b:

(4) $\boldsymbol{P}(\text{ia,ib}|\text{a,b}) = \boldsymbol{P}(\text{ia}|\text{a,rig})\boldsymbol{P}(\text{ib}|\text{b,rig})\boldsymbol{P}(\text{rig}) + \boldsymbol{P}(\text{ia}|\text{a},\overline{\text{rig}})$
$\boldsymbol{P}(\text{ib}|\text{b},\overline{\text{rig}})\boldsymbol{P}(\overline{\text{rig}}).$

If we instantiate for $\bar{\text{a}}$ and b, for a and $\bar{\text{b}}$, or for $\bar{\text{a}}$ and $\bar{\text{b}}$, then the first term of the addition will contain the factors $\boldsymbol{P}(\text{ia}|\bar{\text{a}},\text{rig})$ or $\boldsymbol{P}(\text{ia}|\bar{\text{b}},\text{rig})$. By the probability calculus:

(5) $\boldsymbol{P}(\text{ia}|a,rig) = \boldsymbol{P}(\text{ia}|a,rig,\text{ria})\boldsymbol{P}(\text{ria}|a,rig) +$
$\boldsymbol{P}(\text{ia}|a,rig,\overline{\text{ria}})\boldsymbol{P}(\overline{\text{ria}}|a,rig).$

C1 permits us to read off the following independences by the d-separation criterion:

(6) $ia \perp\!\!\!\perp rig|a,ria$; $ria \perp\!\!\!\perp a|rig$.

We work these independences into (5):

(7) $\boldsymbol{P}(\text{ia}|a,rig) = \boldsymbol{P}(\text{ia}|a,\text{ria})\boldsymbol{P}(\text{ria}|\text{r}ig) + \boldsymbol{P}(\text{ia}|a,\overline{\text{ria}})\boldsymbol{P}(\overline{\text{ria}}\,|rig)$.

Similarly,

(8) $\boldsymbol{P}(\text{ib}|b,rig) = \boldsymbol{P}(\text{ib}|b,\text{rib})\boldsymbol{P}(\text{rib}|rig) + \boldsymbol{P}(\text{ia}|b,\overline{\text{rib}})\boldsymbol{P}(\overline{\text{rib}}\,|rig)$.

We instantiate $\bar{\text{a}}$ and rig into (7) and let (ii) and (iii) reduce each term to 0:

(9) $\boldsymbol{P}(\text{ia}|\bar{\text{a}},\text{rig}) = \boldsymbol{P}(\text{ia}|\bar{\text{a}},\text{ria})\boldsymbol{P}(\text{ria}|\text{rig}) + \boldsymbol{P}(\text{ia}|\bar{\text{a}},\overline{\text{ria}})\boldsymbol{P}(\overline{\text{ria}}\,|\text{rig}) = 0$.

Similarly for instantiating $\bar{\text{b}}$ and rig into (8):

(10) $\boldsymbol{P}(\text{ib}|\bar{\text{b}},\text{rig}) = \boldsymbol{P}(\text{ib}|\bar{\text{b}},\text{ria})\boldsymbol{P}(\text{ria}|\text{rig}) + \boldsymbol{P}(\text{ib}|\bar{\text{b}},\overline{\text{rib}})\boldsymbol{P}(\overline{\text{rib}}|\text{rig}) = 0$.

Hence we may drop the first term in all the remaining instantiations of (3):

(11) $\boldsymbol{P}(\text{ia,ib}|\bar{\text{a}},\text{b}) = \boldsymbol{P}(\text{ia}|\bar{\text{a}},\overline{\text{rig}})\boldsymbol{P}(\text{ib}|\text{b},\overline{\text{rig}})\boldsymbol{P}(\overline{\text{rig}})$;

(12) $\boldsymbol{P}(\text{ia,ib}|\text{a},\bar{\text{b}}) = \boldsymbol{P}(\text{ia}|\text{a},\overline{\text{rig}})\boldsymbol{P}(\text{ib}|\bar{\text{b}},\overline{\text{rig}})\boldsymbol{P}(\overline{\text{rig}})$;

(13) $\boldsymbol{P}(\text{ia,ib}|\bar{\text{a}},\bar{\text{b}}) = \boldsymbol{P}(\text{ia}|\bar{\text{a}},\overline{\text{rig}})\boldsymbol{P}(\text{ib}|\bar{\text{b}},\overline{\text{rig}})\boldsymbol{P}(\overline{\text{rig}})$.

We expand the factors in (11) through (13) by means of (7) and (8). Subsequently, by (i), (iii) and C3, it follows that:

(14) $\boldsymbol{P}(\text{ia,ib}|\bar{\text{a}},\text{b}) = \boldsymbol{P}(\text{ia,ib}|\text{a},\bar{\text{b}}) = \boldsymbol{P}(\text{ia,ib}|\bar{\text{a}},\bar{\text{b}})$.

Furthermore, from (3), (7) and (8) by C3 and C4:

(15) $P(\text{ia,ib}|a,b) = P'(\text{ia,ib}|a,b)$.

Apply Bayes theorem to r. Set $\alpha = P(\text{ia,ib}|\text{a,b})$; $\beta = P(\text{ia,ib}|\bar{\text{a}},\text{b})$; $\text{w} = P(\text{a,b})$; $\text{x} = P(\bar{\text{a}},\text{b})$; $\text{y} = P(\text{a},\bar{\text{b}})$; $\text{z} = P(\bar{\text{a}},\bar{\text{b}})$. Similarly, $\alpha' = P'(\text{ia,ib}|\text{a,b})$ etc. By the identities in (14):

(16) $$\text{r} = 1 \Leftrightarrow \frac{\alpha\text{w}/(\alpha\text{w} + \beta\text{x} + \beta\text{y} + \beta\text{z})}{\alpha'\text{w}'/(\alpha'\text{w}' + \beta'\text{x}' + \beta'\text{y}' + \beta'\text{z}')} = 1.$$

By the identity in (15):

(17) $$\text{r} = 1 \Leftrightarrow \frac{\alpha\text{w}/(\alpha\text{w} + \beta\text{x} + \beta\text{y} + \beta\text{z})}{\alpha\text{w}'/(\alpha\text{w}' + \beta\text{x}' + \beta\text{y}' + \beta\text{z}')} = 1.$$

By the probability calculus, $\text{w} = 1 - (\text{x}+\text{y}+\text{z})$ and $\text{w}' = 1 - (\text{x}'+\text{y}'+\text{z}')$. By C2, w = w'. Hence,

(18) $$\text{r} = 1 \Leftrightarrow \frac{\alpha\text{w}/[(\alpha\text{w} + \beta(1-\text{w})]}{\alpha\text{w}/[\alpha\text{w} + \beta(1-\text{w})]} = 1$$

which completes the proof.

Department of Philosophy LUC BOVENS
University of Colorado at Boulder
CB 232
Boulder, CO 80309
USA
bovens@spot.colorado.edu

ERIK J. OLSSON

Fachgruppe Philosophie
Universitat Konstanz
Postfach 5560 D21
78457 Konstanz
Germany
Erik.Olsson@uni-konstanz.de

REFERENCES

Bender, J. W. (ed.) 1989: *The Current State of the Coherence Theory*. Dordrecht: Kluwer.

Bonjour, L. 1985: *The Structure of Empirical Knowledge*. Cambridge, Mass.: Harvard University Press.

———1999: "The Dialectic of Foundationalism and Coherentism", in Greco, J. and Sosa, E. (eds.) 1999, pp. 117–42.

Bovens, L. and Hartmann, S. (2000a): "Coherence, Belief Expansion and Bayesian Networks", in: C. Baral and M. Truszczynski (eds.) *Proceeding of the 8th International Workshop on Non-Monotonic Reasoning, NMR 2000*. (http://xxx.lanl.gov/abs/cs.AI/0003041)

Bovens, L. and Hartmann, S. (2000b): "Solving the Riddle of Coherence", forthcoming.

Dawid, A. P. 1979: "Conditional Independence in Statistical Theory", *Journal of the Royal Statistical Society* A41: pp. 1–31.

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

Huemer, M. 1997: "Probability and Coherence Justification". *The Southern Journal of Philosophy*, 35, pp. 463–72.

Greco, J. and Sosa, E. (eds.) 1999: *The Blackwell Guide to Epistemology*. Malden, Mass.: Blackwell.

Lewis, C. I. 1946: *An Analysis of Knowledge and Valuation*. LaSalle, Illinois: Open Court.

Neapolitan, R. E. 1990: *Probabilistic Reasoning in Expert Systems*. New York: Wiley.

Olsson, E. J. 2000: "Coherence, Truth and Interpretation", forthcoming.

Pearl, J. 1988: *Probabilistic Reasoning in Intelligent Systems*. San Mateo, Ca: Morgan Kaufmann.

———2000: *Causality*. Cambridge: CUP.

Quine, W. V. O. 1960: *Word and Object*. Cambridge, Mass.: MIT Press.

Spohn, W. 1980: "Stochastic Independence, Causal Independence, and Shieldability", *Journal of Philosophical Logic*, 9: pp. 73–99.

Swain, M. 1989: "Bonjour's Coherence Theory of Justification", in Bender (ed.) 1989, pp. 115–24.

Verma, T. and Pearl, J. 1988: “Causal Networks: Semantics and Expressiveness”, in *Proceedings of the Fourth AAAI Workshop on Uncertainty in Artificial Intelligence*, University of Minnesota, Minneapolis, Minnesota.

———1990: “Equivalence and Synthesis of Causal Models”, in *Proceedings of the 6th Conference on Uncertainty in Artificial Intelligence*, Mountain View, Ca, pp. 220–7. Reprinted in P. Bonissone, M. Henrion, L.N. Kanal, and J.F. Lemmer (eds.) 1991, *Uncertainty in Artificial Intelligence*, vol. 6, Amsterdam: Elsevier, pp. 255–68.