Foundationalism, Probability, and Mutual Support

Shortened Title: Foundationalism and Mutual Support

Abstract

The phenomenon of mutual support presents a specific challenge to the foundationalist epistemologist: Is it possible to model mutual support accurately without using circles of evidential support? We argue that the appearance of loops of support arises from a failure to distinguish different synchronic lines of evidential force. The ban on loops should be clarified to exclude loops within any such line, and basing should be understood as taking place within lines of evidence. Uncertain propositions involved in mutual support relations are conduits to each other of independent evidence originating ultimately in the foundations. We examine several putative examples of benign loops of support and show that, given the distinctions noted, they can be accurately modeled in a foundationalist fashion. We define an evidential “tangle,” a relation among three propositions that appears to require a loop for modeling, and prove that all such tangles are trivial in a sense that precludes modeling them with an evidential circle.
Foundationalism, Probability, and Mutual Support

Foundationalism and the ban on loops

In the long-standing disagreement between foundationalists and their rivals, the foundationalist's insistence that circular reasoning cannot be justificatory plays an important role. This fact is in no small measure a result of the importance to foundationalism of the regress argument. Roughly and briefly, the regress argument says that there must be something that is justified without being inferred from anything else, for otherwise nothing could be inferentially justified. The argument depends on ruling out infinite regresses of reasons and, relatedly, loops of reasons as both being non-justificatory. If all inferential justification is one-directional and if the push “backwards” in search of reasons must come to an end somewhere, then there must be some set of special beliefs that can justify other beliefs without themselves standing in need of inferential support. So argues the foundationalist.

It is easy to see, then, that a challenge to the proposition that loops of support are non-justificatory is *ipso facto* a challenge to foundationalism. For if justification can come through loops in which one proposition is found in its own inferential ancestry, then perhaps foundations are not necessary at all. Perhaps, as the coherentist insists, there is no asymmetry between a privileged set of non-inferentially justified beliefs and all others. Perhaps all of one's beliefs can justify one another in a system that floats free of foundational propositions altogether and stands or falls based on the degree to which the set of beliefs is “mutually explanatory” (Dancy, 2003, p. 207).

Nor need one be a coherentist to see the importance to foundationalism both of the regress argument and of the ban on loops. In *Evidence and Inquiry* Susan Haack advocates a hybrid position that she calls “foundherentism,” ostensibly combining the best and avoiding the worst of both foundationalism and coherentism. But one of the express tenets of foundherentism is that “there need be no privileged class of empirical beliefs justified exclusively by the support of experience, independently of the support of other beliefs” (Haack, 1993, p. 19). Here, too, a rejection of the foundationalist ban on loops features. Haack characterizes “pure foundationalism” as requiring that “a derived belief is justified wholly via the support, direct or indirect, of a basic belief or beliefs” (Haack, 1993, p. 17) and maintains that the possibility of portraying legitimate types of support by way of loops undermines the pure foundationalist assumptions of the regress argument (Haack, 1993, pp. 23-25).

The mutual support objection

Haack argues that “pure” foundationalism is flawed because, lacking loops in its account of evidential relations, it cannot accommodate the sort of mutual support one finds in the cross-hatching entries in a crossword puzzle.

How reasonable one’s confidence is that 4 across is correct depends, *inter alia*, on one’s confidence that 2 down is correct . . . [H]ow reasonable one’s confidence is that 2 down is
correct in turn depends, *inter alia*, on how reasonable one’s confidence is that 4 across is correct (Haack, 1993, p. 86).

Since the “interpenetration of beliefs” illustrated by a crossword puzzle need not involve a vicious circularity, Haack suggests that pure foundationalism must be abandoned. So strongly does Haack associate mutual support with justificational loops that she implies that pure foundationalism, banning loops, denies the epistemic relevance of mutual support altogether; she therefore makes the allowance of mutual support a decisive mark of distinction between what she calls “pure” and “impure” foundationalism (Haack, 1993, pp. 17, 25).

John Post (1996, Post and Turner, 2000, p. 76) comes to a similar conclusion, though by a somewhat different route, claiming that the mutual support evident in the probabilistic relations of theory and supporting data requires placing a proposition in its own inferential ancestry. A well-attested piece of data such as Neptune's position at some particular time supports Newton's theory, while on the other hand Newton's theory entails that Neptune will appear in that position. Insofar as Newton's theory is itself well-supported independently of the position of Neptune, one will in fact expect to find Neptune in the stated position. This, argues Post, is an example of non-vicious circularity, since the propositions in question are mutually justificatory and, Post says, the proposition about Neptune's position must appear in its own inferential ancestry, but no proposition is justifying itself. Hence, it must be legitimate to model support using inferential loops, and the foundationalist's regress argument (with its conclusion that foundations are necessary) must be seriously flawed (Post and Turner, 2000, p. 67).

The suggestion that foundationalism is mistaken because it rules out loops in evidential relations is quite remarkable. If it were successful, the criticism would amount to at least a limited vindication of circular reasoning. We do advocate the position that all justificatory force comes ultimately from foundational premises, and this position requires that we reject loops of support. These arguments therefore place on us the onus of showing how we can accommodate so intuitively legitimate a justificatory structure as that which holds between mutually supportive propositions. The challenge, then, is to model the phenomenon of mutual support without violating the principle that circular reasoning is non-justificatory.

*A flawed attempt at a solution*

In an earlier attempt to analyze mutual support from a foundationalist perspective, we argued that a foundationalist should understand it diachronically (identifying references omitted). According to this analysis, it is not the case that all evidential relations can be exploited simultaneously. To avoid loops of support, a rational agent must choose to prioritize some of his beliefs rather than others; beliefs that are not prioritized will not receive the maximum support possible given total evidence if some of this support comes from beliefs that they are in turn supporting. At a different time, however, and without the introduction of new evidence, beliefs previously treated as purely supportive could be taken as conclusions on which maximal available support was brought to bear.

There is *something* right about this analysis. The fact that two beliefs are positively correlated means that either of them could be taken, under appropriate circumstances, to supply some
reason for the other; and it obviously does not follow from this that we would be justified in believing any two propositions with a sufficiently strong positive correlation. But to say this amounts, upon examination, only to an admission of the phenomenon of positive correlation together with a reiteration of the ban on circular reasoning. It does not tell us how to model mutual support.

The diachronic approach, unfortunately, does not really solve the synchronic problem as to what a rational agent’s degrees of confidence will be. On a very modest interpretation, this approach considers only what the subject’s conditional probabilities will be when he is temporarily “setting aside” some of his evidence and considering the probability of some proposition conditional on the remaining evidence. But this does not tell us what his degrees of confidence should actually be on total evidence.

More radically, the diachronic approach and the concept of prioritizing suggest that a rational subject may at one time literally and even deliberately ignore some of his relevant evidence for some of the propositions in his distribution and, at a different time, take that evidence into account for those propositions but ignore some evidence pertinent to others. That is, if a subject happens to be aware of evidential relevance that goes from A to B and also from B to A, he must pretend at one time that he is not aware of the evidential connection in one direction, fixing his credibilities accordingly, and, if he re-prioritizes, pretend that he is not aware of the connection in the other direction and have, in consequence, a different set of credibilities. But this answers the synchronic question at too high a price, for it violates the requirement that the subject be perfectly rational at the synchronic moment. We are strongly committed to what we would call the objectivity constraint on rational belief:

Disagreements regarding the probability of any proposition are due either to differences in the relevant evidence available to the disagreeing parties or to specific inferential failures on the part of at least one of the disputants.

But if two subjects with the very same evidence can be equally rational despite assigning different probabilities to their inferred beliefs in virtue of prioritizing their beliefs differently, this condition does not hold.

Hence, the diachronic attempt to solve the question of mutual support either evades the question of actual synchronic credibilities on total evidence or deteriorates into subjectivism. The problem of mutual support, then, requires a different solution.

Response to the mutual support objection

To answer the mutual support objection in a satisfactory way, the foundationalist needs to construct a model in which the impact of foundational evidence on higher level beliefs in mutual support scenarios is displayed without generating circularity or violating the objectivity constraint. The most difficult part of this, it turns out, is the clarification of the notions of support and circularity. Both concepts can be clarified by reference to Figures 1 and 2.
Appearance of Circularity:
H₁ and H₂ seem to be based on each other

Figure 1

In Figure 1, four unproblematic foundational beliefs provide support for H₁ and H₂. The beliefs F₁ and Fₐ support H₁ directly: their impact on H₁, both individually and jointly, is positive independent of whether H₂ is true. The higher level beliefs H₁ and H₂ are mutually relevant, but ±H₁ (the assertion or the negation of H₁) screens off F₁ and Fₐ from H₂. That is, with either H₁ or ¬H₁ given on the background, F₁ and Fₐ are probabilistically irrelevant to H₂. Formally, P(H₂ | H₁ & F₁ & Fₐ) = P(H₂ | H₁ & F₁) = P(H₂ | H₁ & Fₐ) = P(H₂ | H₁), and the same mutatis mutandis for ¬H₁. The situation is symmetrical with respect to F₂, F₁, and H₂: they provide, individually and jointly, direct positive support for H₂, but ±H₂ screens off their impact on H₁.

The arrows in Figure 1 do indeed give an impression that there is something circular going on. The eye travels from f₁ up through H₁ and across to H₂, but then doubling back we come back to H₁ again. Moreover, we have stipulated that H₁ and H₂ are mutually relevant, so the arrow going from left to right and the arrow from right to left seem equally legitimate. (After all, when P(H₁|H₂) > P(H₁), P(H₂|H₁) > P(H₂).) How can we dispel the appearance of circularity here?

The central idea of the solution is that simple conditionalizing on the foundational beliefs propagates through the higher levels of the distribution in what are, formally speaking, Jeffrey conditionalizations on intermediate nodes. By attending to the screening condition, we are in a position to refine the notion of support: instead of saying that H₂ supports H₁ simpliciter, we can say more accurately that H₂ is a conduit through which F₂ and Fₐ support H₁, and similarly that H₁ is a conduit by which F₁ and Fₐ support H₂. The foundationalist’s restriction on circularity is thereby similarly refined: no proposition may support itself, in the sense that no line of support may originate or pass through a belief node and also return to it.

Suppose that two of the foundational beliefs – for simplicity, Fₐ and F₂ – are fixed at 1. What happens as we acquire F₁ and F₂? We can, if we like, model this by acquiring first one and then the other; since the shift is induced by the acquisition of certainties at the foundational level, order is immaterial. Due to the screening off conditions, a change in H₁ from one intermediate
value to another is propagated to $H_2$ via a shift that can be modeled by the formula introduced by Richard Jeffrey and known as the Jeffrey conditioning formula or the formula for probability kinematics:

$$P_{\text{new}}(H_2) = P_{\text{old}}(H_2|H_1) P_{\text{new}}(H_1) + P_{\text{old}}(H_2|\neg H_1) P_{\text{new}}(\neg H_1)$$

The “old” distribution here refers to the distribution where $F_A$ and $F_B$ are in the background but where $F_1$ has not been acquired. The “new” distribution is the distribution where we have acquired $F_1$ as a new certainty. As has been noted by more than one person (e.g., Pearl, 1988, p. 64), the screening conditions described here guarantee that the technical requirements for using Jeffrey's formula in this way, known as the rigidity conditions, are satisfied. And although we could express the shift in the probability of $H_2$ in terms of simple conditioning on the newly acquired certainty, $F_1$, doing so would show only a “direct” relation between $H_2$ and $F_1$, leaving $H_1$ out of the picture. Portraying the shift in terms of simple conditioning alone, while not technically incorrect, fails to illustrate the relationship we want to draw attention to between $H_2$ and $H_1$, the two propositions involved in a mutual support relationship.

Something exactly similar can be said regarding the acquisition of $F_2$: Its acquisition, as well, can be modeled as a shift in $H_1$ using Jeffrey's formula “on” $H_2$. The acquisition of $F_2$ induces a new probability in $H_2$, and the new probability of $H_1$ can be calculated using the new intermediate probability of $H_2$. Here the “old” distribution includes $F_1$ as given (which we introduced in the step above), but not $F_2$. The “new” distribution here includes the newly-acquired $F_2$ as given. Again, order of updates does not matter since the change from the old to the new distribution is induced by the acquisition of certainties.

$$P_{\text{new}}(H_1) = P_{\text{old}}(H_1|H_2) P_{\text{new}}(H_2) + P_{\text{old}}(H_1|\neg H_2) P_{\text{new}}(\neg H_2)$$

The ultimate source of the change in the probabilities is the acquisition of foundational information. But for the intermediate propositions the Jeffrey conditioning formula shows clearly how the probabilities are coherently redistributed as new evidence is acquired.

It is perhaps only fair to note that this use of Jeffrey's formula is diametrically opposed to Jeffrey's original intention in introducing the formula. He meant it to be a substitute for simple conditioning on propositions with probability 1. Jeffrey's idea was that the change in the probability of uncertain evidence (in this case, of $H_1$ or $H_2$) should be a kind of surd, brought about by a passage of experience (at least, so it is in Jeffrey's examples), but not brought about by the acquisition of some new certain proposition as evidence, an option he expressly rejected (see, for example, Jeffrey 1990, pp. 164-8, 1992, 135-6). However, the mathematical formula is just that: a formula, and a consequence (under the stated rigidity conditions) of the Theorem on Total Probability. As such it is available for use by those who do not share Jeffrey's severe doubts about the availability of foundational propositional evidence. And it is extremely useful in allowing the foundationalist to model the way that a shift in the probability of one uncertain proposition is related to that of another, as is the case in mutual support.

The upshot is that we can show, by means of two different diachronic paths to a single final coherent distribution, how each of the propositions $H_1$ and $H_2$ can channel epistemic support to
the other without any incoherence, circularity, or violation of the objectivity constraint. This does not mean that the reasoning must be diachronic. Rather, the diachronic modeling merely shows that the evidence is indeed making a difference, and the fact that order is immaterial supports the contention that we are by this means understanding the synchronic support provided by both pieces of evidence in the final distribution. This solves the problem of mutual support.

All of this is well illustrated in Figure 2, where the black arrows of the first figure have been replaced by various styles of arrows that give us more insight into the underlying support relations. Critically, there is nowhere a cyclical path of the same style. This permits a visual representation not only of the positive correlation of \(H_1\) and \(H_2\), which is of course a symmetric relation and gives us the “mutual” in mutual support, but also of the channeling of epistemic support from more fundamental premises through \(H_1\) to \(H_2\) and \textit{vice versa}, without circularity.

![Mutual Support without Circularity:](image)

\[
\text{Mutual Support without Circularity:}
\text{No line of support goes both from } H_1 \text{ to } H_2 \text{ and } \textit{vice versa}
\]

Figure 2

\textit{The basing relation}

Perhaps the most intriguing implications of this analysis of mutual support concern the concept of basing. If we conjoin an unqualified and unclarified statement that basing is transitive with an acknowledgment of synchronic mutual support, we will be forced to endorse loops of epistemic support, contrary to the norms of foundationalism. \textit{Prima facie}, if \(A\) is based on \(B\) and \(B\) is based on \(A\), then \(A\) is based on itself – the picture suggested by Figure 1.

What is needed, then, is a more detailed and sophisticated notion of basing. The foundationalist can and should insist that basing be thought of as taking place only within a line of support. On this conception, some propositions are based directly on the foundations while others are supported by way of intermediate-valued “conduits” that channel the epistemic force of the foundations to other propositions. The line of support passes from the foundations, through conduit nodes (if there are any) to some other proposition or propositions in the distribution. Understanding the epistemic underpinnings of the subject’s rational credibility for a proposition is thus not simply a matter of listing all of the propositions that appear anywhere in any
probabilistic relation either to it or to any of its premises. We must follow the lines of support.\(^3\)

The notion of basing a belief on a conduit node requires some explication. First, a premise in some given subject’s cognitive structure must be treated as such by the subject. No belief, foundational or non-foundational, can be a premise for a particular subject merely as a result of probabilistic facts, if he is not making use of those probabilistic facts inferentially. He must be basing B on A for A to be a premise for B, and it may well be that this (in one sense) psychological aspect of basing is *sui generis*, not admitting either of a purely causal or of a purely probabilistic analysis. And we are not inclined to capture this idea of basing by saying that the subject actually knows or must know metalevel facts about evidential and probabilistic relations, though we would say that a logically omniscient “upgraded version” of the subject would indeed know these facts.\(^4\)

The more tractable aspect of what it means for a premise to be a conduit is the evidential and probabilistic one. Here we are indebted to the notion of a basis articulated by Hawthorne:

An evidence basis is a partition – i.e., a mutually exclusive and exhaustive set of sentences. The evidential impact of an experience or observation \(e\) is supposed to be completely captured by the influence it has on the belief strengths of sentences of its evidence basis....The evidence basis captures the evidential import of \(e\) through its ability to “screen off” the rest of the agent’s beliefs from direct influence by \(e\) (Hawthorne, 2004, p. 93).

We would add, what is entirely compatible with Hawthorne’s explication of a basis, that the propositions of the basis may have to be complex, consisting (for example) of the negations and assertions of more simple component propositions in all their permutations, such as \{(A & B), (A & ~B), (~A & B), (~A & ~B)\}. Moreover, we should note that the propositions in such a basis must be evidentially relevant to the propositions to which they channel evidence and that the foundational evidence thus channeled should be relevant to the higher-level propositions; otherwise the concept of a basis can be trivialized.\(^5\) Furthermore, as Hawthorne has emphasized, the basis need not screen off foundational evidence from all other propositions in the distribution.\(^6\) This means that to say that some partition is a basis for some foundational evidence is elliptical: A basis should be understood relative both to a piece of given evidence, on the one hand, and to some particular other proposition or propositions in the distribution, on the other, even if it does not screen the given evidence from every other proposition.\(^7\)

Finally, there may be more than one screening conduit “node” – more than one partition – through which a line of epistemic force passes on its way from the foundations to some other uncertain proposition we happen to be considering. That is, some piece of foundational evidence may have an impact on some higher-level proposition D by way of a line that passes “through” node A, then through node B, then through node C. We illustrate the tracing of such an evidential line when considering the example from Peter Vranas, below.

Once we understand that premises can be conduits of given evidence, it becomes clear that the actual probability of \(H_1\) at \(t\), conditional on *all* evidence pertinent to \(H_1\), is not the relevant point to consider in understanding how or whether it is supporting \(H_2\).\(^8\) It is not as though it can be supportive only if its probability meets some criterion such as having probability of greater than .5 on total evidence. For it could be channeling positively relevant foundational evidence to
H₂ even if its own probability is low. And some of the evidence pertinent to and supporting H₁ may be channeled by H₂ itself as a conduit. If this is the case, it is doubly important that we not think of the probability of H₂ on total evidence as critical to its being or not being a premise for H₂. Nor will we come to understand the role H₁ plays with respect to H₂ in one’s cognitive economy better simply by calculating \( P(H_2) \) in terms of a partition \{H₁, ~H₁\}. Any two propositions in the distribution can be “related” by such a calculation. This does not tell us whether or how either of them is acting as a premise for the other.

The notion of a conduit is not arcane to a foundationalist, particularly not to a strong foundationalist who holds that the foundations must be certainties; he has been tacitly using it all along, even if he has not realized its importance for mutual support. A foundational premise, which on his theory has at time t probability 1, is a premise in a sense that requires no explication or qualification. But strong foundationalists have never conceded that non-certain propositions are originators of evidential force. Hence, propositions that are not known with certainty, whether rationally believed or not, are premises in a sense that requires a more nuanced explication to take into account the fact that they are not foundational. Any intermediate-valued proposition that a strong foundationalist is inclined informally to call a ‘premise’ for some other proposition is, formally, either a conduit channeling the force of some foundational evidence to that proposition or else a simpler sub-component of the complex members of a conduit node which channels foundational evidence to that proposition. In terms of Figure 2, H₁ and H₂ are conduits of support to each other, and it is for exactly this reason that there is no circularity. Each supports the other only in the sense that and insofar as it channels evidential support to the other unidirectionally from elsewhere in the distribution and, ultimately, from the foundations. H₁ is based on H₂ only within the lines of support that run from \( F_B \) and \( F_2 \) upwards, and H₂ is based on H₁ only within the lines of support that run from \( F_A \) and \( F_1 \) upwards.

This notion – that basing occurs only within lines of evidence – clarifies the issue of transitivity. Basing is transitive, but that is because it occurs only within evidential lines and must always be understood relative to such a line. And lines of support are themselves acyclic. Within the line of support from \( F_B \) and \( F_2 \), H₁ is based on H₂, H₂ is based on \( F_B \) and \( F_2 \), and H₁ is based on \( F_B \) and \( F_2 \) via their impact on H₂ as a conduit. There is here neither any loop nor any counterexample to transitivity. And the same is true \textit{mutatis mutandis} for the lines that run from \( F_1 \) and \( F_A \) to H₂ via H₁.

\textit{The trouble with loops}

Here, however, it may seem that we have smuggled in the assumption of foundationalism by stipulating that lines of support are acyclic. Why, the anti-foundationalist might ask, could they not contain loops?

First, we should stress that the challenge was for a foundationalist to model mutual support without violating his own norms, so it is not relevant to point out that, in so doing, we have kept in place a ban on loops of support. If loops are not needed for modeling mutual support, anti-foundationalists will have to find some other reasons for questioning their exclusion.

But we need not stop at asking the anti-foundationalist to drop his objection. Armed with a better understanding of basing and its transitivity, we can also shed light on the foundationalist’s
prohibition against loops. For what could it mean to speak of a “loop of support,” or more generally of a looped branch in an evidence tree? Evidently it would mean that the impact of some evidence F on some particular belief B would be felt more than once, that the subject’s noetic structure included something like an epistemic ratchet that could be jiggled repeatedly and indefinitely. But the impact of the evidence is exhausted by its being taken into account once, either directly or via Jeffrey conditionalization on an intermediate basis that screens F from B. Any attempt to take it into account a second time would either induce incoherence or indicate that the first “conditionalization” had not exhausted the relevance of F to B – in which case, of course, it was not a legitimate conditionalization in the first place.

We need to make it clear that no one is explicitly advocating ad infinitum self-support. In fact, Haack insists upon independent support for each of the two mutually supporting propositions (Haack, 1993, p. 86) as does Post (2000, p. 76). But if real self-support, real bootstrapping, is not in view, then what is the alternative to foundationalism supposed to be? And where is the challenge in the phenomenon of mutual support?

Perhaps the whole appeal of the mutual support objection can now be reduced to one of representation, to a mental image of arrows going from A to B and vice versa. But the strong foundationalist should never have conceded in the first place that support comes from uncertain evidence A to uncertain proposition B, as if the evidential force originated in A. He should, rather, have insisted all along that support goes through uncertain A to uncertain B. Once this is clearly understood, the confused image of loops of support can be resolved into a clearer one of unidirectional lines of support passing through two different propositions, and the mutual support objection to foundationalism collapses.

It might be asked here whether our objection to loops is chiefly a terminological one. Could we not call the phenomenon of mutual support a case of “benign circularity” or a “benign loop” while explaining ourselves carefully in the terms we have outlined here? Why should the foundationalist object to such terminology?

The simplest response is that in view of the foregoing analysis, such terminology blurs the underlying epistemic structure by limiting our analytical resolving power. We might well turn the question around: Why, now that we have shown how the lines of force in cases of mutual support should be probabilistically disentangled into one-directional strands, would the non-foundationist hold out for the imprecise and confusing terminology of “benign circularity”?

Here Haack's characterization of “pure” and “impure” foundationalism is helpful in showing both why the pure foundationalist should reject loops of support and why others have been motivated to accept such a representation. As we have already noted, Haack characterizes pure foundationalism as requiring that justified beliefs receive all of their support directly or indirectly from basic beliefs. But in her taxonomy, impure foundationalism and a fortiori Haack's own foundherentism do not:

[A]ccording to the impure version, derived beliefs are always justified at least in part by means of the support of basic beliefs, but the possibility is allowed that they may get part of their justification by means of mutual support among themselves (Haack 1993, p. 17, emphasis in original).
So those who reject pure foundationalism hold that there is some sort of extra evidential force that arises from mutual support among non-foundational beliefs *above and beyond* support that comes ultimately from the foundations. A looped branch in an evidence tree, then, indicates the existence of evidential force that does not come one-directionally, directly or indirectly, from the foundations. But this is precisely what we deny. The ban on loops thus signals adherence to a principle that lies at the heart of pure foundationalism: There is *no* epistemic force generated in support relations apart from the evidential impact of the foundations. We have modeled mutual support in a way that is in keeping with this principle, showing that no extra probabilistic force arises out of the inter-relatedness of non-foundational beliefs alone. All that is required to see that all force is coming from the foundations is the patient work of teasing out the lines of evidence that pass from the foundations through various non-foundational beliefs to others. Mutual support is important and lends great strength to our beliefs precisely because it represents the convergence of several different one-directional lines of evidence upon each of the non-foundational propositions involved. Hence it would be not merely infelicitous but seriously misleading to characterize mutual support in terms of loops.

*Where are the benign loops?*

But has the foundationalist really addressed all of the cases that the anti-foundationalist has in mind? Are there “benign loops” to which our analysis is inadequate? Putative examples of benign loops are not easy to find, and detailed, probabilistically explicated examples are even rarer. Haack’s example of the crossword puzzle can be accommodated in the way we have already indicated, and Haack’s own insistence that the propositions bring independent evidence to bear on each other is entirely in keeping with our own analysis.
Suppose that, as in the crossword puzzle Haack gives, 4 across is a four-letter word and the clue is “She's a jewel.” (Haack, 1993, p. 85. See figure 3.) Let us assume a dictionary from which words can be taken which gives us a very low (say, 1/20,000) background probability that any given word in the puzzle will be 'ruby'. The evidence of the word's length and the meaning of the clue, of course, raise this probability considerably. Moreover, the foundationalist is entirely free to characterize this independent evidence that 4 across is 'ruby' as coming ultimately from the puzzle-solver's memory-like experiences about the meanings of the words in the clue, his sensory experiences of the appearances on the page that show him the four letter boxes, and so forth. If 4 across is 'ruby', this of course entails that its first letter is R. That 4 across and 2 down intersect at the first letter of 4 across and the second letter of 2 down may be taken as background evidence affecting both propositions. The clue and length for 4 down indirectly raise the credibility of the proposition that 2 down is 'irate', since on normal background evidence a proposition or set of propositions about 4 across will screen the impact of the direct evidence about 4 across from the proposition that 2 down is 'irate'. Similarly, the clue and the number of letters in 2 down constitute independent lines of evidence that affect directly the proposition that 2 down is 'irate' and indirectly (through a proposition or set of propositions about the specific word at 2 down) the proposition that 4 across is 'ruby'. So nodes concerning the correct word for each of the two entries stand in the places of H₁ and H₂ in our Figure 2, above, channeling to one another the impact of independent evidence.

In the same way, the mutual support between theory and data mentioned by Post and Turner (2000, p. 76) can be analyzed according to the model of Figure 2. The proposition that Newton's
theory is true channels the force of independent evidence that supports that theory (e.g., that the orbit of the moon is predicted accurately to within the error of measurement by an inverse square gravitational attraction) to the claim that the data are indeed as the theory say they should be – specifically, that Neptune is found in a particular position. And in fact, if we have plenty of other evidence that Newton's theory is correct, we will actually expect to find Neptune where that theory predicts it to be. Such lines of support for Newton's theory go in one direction – to Newton's theory and “through” it to the proposition about the orbit of Neptune, which they support indirectly. Independent evidence (say, an observation report) that the planet is in that position provides a different line of support going in the opposite direction, passing through the proposition about Neptune's position to Newton's theory. There is no need to represent this mutual support with a loop.

The most careful and probabilistically sophisticated attempt to exhibit a benign loop of which we are aware is the following example, which we owe to Peter Vranas. Suppose that there are three coins, A, B, and C. A is a fair coin, whereas B and C are both biased towards heads, though not altogether two-headed. (For concreteness, we may suppose that B is biased 60-40 in favor of heads and C is biased 70-30 in favor of heads.) Now let X be the claim that coin A falls heads. Let Y be the claim that coins A and B fall the same, whether heads or tails. Let W be the claim that coins A and C fall the same, whether heads or tails, and let Z be the complex claim that either all three coins fall heads or exactly two fall tails. When all of these probabilities are in place, we have the following situation: Each of these four propositions is positively relevant to each of the others, and ±Y screens X from Z, ±Z screens Y from W, ± W screens Z from X, and ± X screens W from Y.

As Figure 4 illustrates, this creates a ring of nodes for a possible evidence tree. Any three adjacent nodes can be related by the fact that the middle one screens the first from the third and that the first is positively relevant to the third. This presents the foundationalist with the challenge to say where the line or lines of evidence should go. At first blush, it appears that choices will be arbitrary and that some of these possible evidential connections, if put together, will result in a loop.

But where is the evidence coming from? It is important to bear in mind that the foundationalist will always model lines of support as originating in the foundational evidence.
Merely pointing out that propositions bear a relation of mutual positive probabilistic relevance to each other does not amount even to a plausible challenge to the foundationalist's model. To a strong foundationalist, a probability distribution is not a surd. The propositions do not simply have these probabilities and relations in and of themselves. The first thing to do is to seek the origin of any line of support. As it stands, this ring does not represent mutual support or any other kind of support, because its arrows merely represent probabilistic relevance, not the flow of evidential force.

It is important, though, that we not ruin the apparent challenge of the example in the very act of bringing in some outside evidence at probability 1. If, for example, we envisage some proposition F which, when given at probability 1, raises the probability of node X (that coin A falls heads) to .7, the screening relations that make the example so interesting no longer obtain. Specifically, ±Z no longer screens Y from W. What we must do to show the flow of a line of evidence while retaining the challenging properties of the example is to make the outside evidence induce the distribution itself. Let us, then, set coin A as tail-biased in an old distribution and let evidence F, when given at probability 1, induce the challenging distribution by making coin A fair – by bringing the probability of X up to .5. Now, in the new distribution, all the screening and relevance relations just described hold, but we have a line of evidence coming in whose flow we can trace through the rest of the sub-algebra.

Setting F to probability 1 raises the probability of all the rest of the nodes. This makes sense semantically. For if coin A – named in X – was tail-biased before, but coins B and C (the different coins named in Y and W) are heads-biased, it will of course be more probable that coin A will fall the same as either of them when it is fair. And even the complex and unusual Z is understandably more probable in the new than in the old distribution, for now there is a greater probability of the disjunct in which all three coins fall heads. So how is the evidence from F routed through the other nodes?

The first thing to notice is that ±X screens F from everything else in the “ring.” So the line of evidence should be drawn rising from F through X and not going to any of the other nodes by any other route. At this point, the evidence tree will fork, with arrows rising from X to each of Y and W, for in the new distribution induced by F neither Y nor Z screens X from W, and neither W
nor Z screens X from Y. (See Figure 5.) This, too, makes sense semantically and informally. For example, if coin A falls heads, this raises the probability that it falls the same as heads-biased coin C, and it does so whether or not we know that it falls the same as coin B.

The branching move whereby the evidence from F rises to Y and W is very important, because it removes the apparent challenge posed by the fact that ±X screens W from Y. There should be no arrow coming from W to Y or vice versa, because the evidential support is coming upwards from F, through X, to each of W and Y. No independent support is being channeled through W to the rest of the sub-algebra – to Y, for example. Nor is any coming through Y to W. Thus the apparent ring is “snipped” at a critical point, despite the existence of the screening relations and the positive relevance among the propositions.

Similarly, and crucially, the fact that ±Z screens Y from W and that Y is positively relevant to W does not mean that arrows in the evidence line should go from Y through Z to W or vice versa. That aspect of the “ring” does not form a portion of the line of evidence from F. For ±Z does not screen either F or X from W – the evidence in this line passes from F only through X to W, not through Z. There is no shift that follows Jeffrey’s formula on Z to W when F is introduced.\(^\text{16}\)

So we see that neither probabilistic mutual relevance nor even local screening relations (among some adjacent three nodes) create a line of evidence. At the most, when three nodes are related by screening and relevance, as described above, such a situation shows a place where a line of support might pass if some evidence is given in that distribution and flows one-directionally through that group. But the important thing for tracing some given actual line is the way in which screening helps to show the flow of evidence from the foundations in that line.

Now we come to an interesting fact. In the distribution induced by F, not one but two nodes screen X from Z. PROB (Z| ±Y & X) = PROB (Z| ±Y), and the same is true of ±W. Each of these nodes also screens F from Z, so the change in the probability of Z can be modeled either by way of the change in Y or in W – as coming “through” either of these two nodes. We seem to have an embarrassment of riches when it comes to ways in which we could show the flow of the
evidence rising from F through X to Z. Whether one calculates \( \text{PROB}(Z) \) by Jeffrey's rule using W or using Y, \( \text{PROB}(Z) \) comes to the same thing – .5. This is no accident, as otherwise the distribution would be incoherent, yielding two different probabilities for Z. And the symmetry between W and Y (each represents the claim that the same fair coin falls the same as a heads-biased coin) makes it not entirely surprising that their relations to X and to Z should be symmetrical. Semantically, these screening relations arise from the fact that, if it is given only that, e.g., coin A falls the same as coin C, the truth or falsehood of Z will be determined entirely by the fall of coin B. In that particular scenario, Z is true if and only if coin B falls heads. But how should the evidence tree be drawn?

We propose that in this sort of case, the foundational evidence gives rise to a set of two possible and probabilistically equivalent evidence trees showing evidence rising to Z, as represented in Figure 6.

![Figure 6](image)

Since the force of F propagates only once, a single evidence tree should not show both lines, as this would give the false impression that “double-dipping” is occurring. It is then a matter of indifference whether one chooses to say that Z is based on F through W and through X or that Z is based on F through Y and through X. The probabilistic outcome is the same, so this is in no sense a return to the old “prioritizing” model in which the subject ignores evidential relevance or chooses from non-equivalent evidence trees on the basis of his interests. Indeed, a fully aware subject would be able to see that the two ways of construing support are equivalent. And, most importantly, in neither of the two evidence trees is there any loop. The absence of loops through Z should be no surprise given our analysis thus far; Z is at the apex of this particular group as influenced by this particular piece of outside evidence, because Z does not channel the force of that evidence to any other node in the group.

What then, in terms of screening and relevance, does the prohibition on loops in the evidence tree rule out? There is one sort of hypothetical challenge to the modeling techniques described here that might appear to create a serious problem. Suppose that there are three nodes of an evidence tree related in such a way that A is relevant both to B and to C, that ±B screens A from C and that ±C screens A from B. Now suppose that A is given evidence. It would seem that the impact of A on B and C cannot be modeled without the use of a loop, since by the conventions
we have been using so far, we would apparently be forced to say that A influences B through C and also that A influences C through B. But the proof given in the appendix shows that all such probabilistic situations (which we dub ‘tangles’) are trivial in the sense that the B and C nodes can be collapsed into a single node influenced by A, so that it is neither necessary nor correct to model A’s influence on C by showing an arrow passing through B or vice versa. The prohibition on loops rules out non-trivial tangles, but all tangles are provably trivial.

**Conclusion**

Inferences involving mutual support are, of course, non-deductive. It is therefore not surprising that the details of probabilistic reasoning should be relevant to understanding mutual support. But such an understanding is just what has been lacking in discussions of mutual support among traditional epistemologists. Evidence trees (or bushes) sprouting arrows in all directions are a poor substitute for a detailed and accurate model of the propagation of evidence induced by changes in the foundations. And once these probabilistic features are taken into account, the challenge of mutual support to foundationalism evaporates.
Appendix: All Tangles Are Trivial

Part I: Proof that all tangles are trivial

Definitions

*Tangle:* A set of propositions \{A, B, C\}, all with regular (non-extremal) probabilities, constitutes a tangle iff:

1*. A is relevant to both B and C,
2*. ±B screens A from C, and
3*. ±C screens A from B.

The tangle is *trivial* iff:

4*. \(P(C|B) = 0\) (equivalently, \(P(B|C) = 0\)) or \(P(C|B) = P(B|C) = 1\).

Triviality, then, arises when the propositions B and C exclude each other (in the sense that the probability of each given the other is 0) or each entails the other (in the sense that the probability of each given the other is 1).

Conditions

Let \{A, B, C\} constitute a tangle. Then:

1. \(P(B|A) \neq P(B)\) [from 1*]
2. \(P(C|A) \neq P(C)\) [from 1*]
3. \(P(C|B&A) = P(C|B)\) [from 2*]
4. \(P(C|\neg B&A) = P(C|\neg B)\) [from 2*]
5. \(P(B|C&A) = P(B|C)\) [from 3*]
6. \(P(B|\neg C&A) = P(B|\neg C)\) [from 3*]

Suppose, further, that the tangle is non-trivial. Then, without loss of generality:

7. \(0 < P(C|B) < 1\)

To simplify notation, let \(P^*\) be a probability distribution arrived at from P by simple conditioning on A, so that \(P^*(-) = P(-|A)\). Then we can rewrite these seven conditions thus:

1. \(P^*(B) \neq P(B)\)
2. \(P^*(C) \neq P(C)\)
3. \(P^*(C|B) = P(C|B)\)
4. \(P^*(C|\neg B) = P(C|\neg B)\)
5. \(P^*(B|C) = P(B|C)\)
6. \(P^*(B|\neg C) = P(B|\neg C)\)
7. \(0 < P(C|B) < 1\)

Lemma 1: \(P^*(B), P^*(C) > 0\).

\(P(B) > 0\), by the definition of a regular probability function. Suppose \(P^*(B) = 0\). Then \(P^*(C|B)\) is undefined. But since \(P(B) > 0\), \(P(C|B)\) is defined. But by 3, \(P^*(C|B) = P(C|B)\). Hence both conditional probabilities must be well defined. Therefore, \(P^*(B) > 0\). Exactly similar reasoning establishes that \(P^*(C) > 0\).
From 1, without loss of generality, \( P^*(B) = kP(B) \), for some \( k \neq 1 \); \( P^*(B)/P(B) = k \). Hence, \( k > 0 \).

**Lemma 2:** \( P^*(B\&C)/P(B\&C) = k \).

**Proof of Lemma 2**

\[
\begin{align*}
P^*(C|B) &= P(C|B) \quad [3] \\
P^*(C&B)/P^*(B) &= P(C&B)/P(B) \quad \text{[def of conditional probability; P(B) > 0 by regularity; P^*(B) > 0 by Lemma 1]} \\
P^*(C&B)/P(C&B) &= P^*(B)/P(B) \quad \text{[multiplying both sides by P^*(B)/P(C&B); P(C&B) > 0 by 7]} \\
P^*(C&B)/P(C&B) &= k \quad [substitution]
\end{align*}
\]

**Lemma 3:** \( P^*(C)/P(C) = k \).

**Proof of Lemma 3**

\[
\begin{align*}
P^*(B|C) &= P(B|C) \quad [5] \\
P^*(C&B)/P^*(C) &= P(C&B)/P(C) \quad \text{[def of conditional probability; reg. of P(C); Lem 1]} \\
P^*(C&B)/P(C&B) &= P^*(C)/P(C) \quad \text{[multiplying both sides by P^*(C)/P(C&B); 7]} \\
P^*(C)/P(C) &= k \quad [substitution]
\end{align*}
\]

**Lemma 4:** \( P^*(B\&~C)/P(B\&~C) = k \).

**Proof of Lemma 4**

\[
\begin{align*}
P(B) &= P(B\&~C) + P(B\&C) \quad \text{[def.]} \\
P^*(B) &= P^*(B\&~C) + P^*(B\&C) \quad \text{[def.]} \\
k P(B) &= P^*(B\&~C) + k P(B\&C) \quad \text{[substitution; P(B\&C) = P(C&B)]} \\
P(B) &= 1/k P^*(B\&~C) + P(B\&C) \quad \text{[dividing through by k]} \\
1/k P^*(B\&~C) &= P(B\&~C) \quad \text{[substituting and subtracting P(B\&C)]} \\
P^*(B\&~C) &= k P(B\&~C) \quad \text{[multiplying through by k]} \\
P(B\&~C) > 0 \quad \text{[from regularity of P(B) and 7]}
\end{align*}
\]

Why? Because \( P(B\&~C) = P(B)P(~C|B) = P(B) [1 – P(C|B)] \). Regularity guarantees that \( P(B) > 0 \), and condition 7 guarantees that \( [1 – P(C|B)] > 0 \).

\[
P^*(B\&~C)/P(B\&~C) = k \quad [dividing by P(B\&~C)]
\]

**Lemma 5:** \( P^*(~C)/P(~C) \neq k \).

**Proof of Lemma 5**

We know that \( k \neq 1 \) [from 1]; therefore either \( k > 1 \) or \( k < 1 \). Suppose \( k > 1 \). Then \( P^*(C) > P(C) \); but in that case \( P^*(~C) < P(~C) \), so \( P^*(~C)/P(~C) < 1 \), in which case \( P^*(~C)/P(~C) \neq k \).

Suppose \( k < 1 \). Then \( P^*(C) > P(C) \); but in that case \( P^*(~C) > P(~C) \), so \( P^*(~C)/P(~C) \neq k \).
> 1, in which case \( P^*(\neg C)/P(\neg C) \neq k \). Therefore, \( P^*(\neg C)/P(\neg C) \neq k \).

**Theorem:** Conditions 1-7 are not simultaneously satisfiable.

**Proof of Theorem**

\[
P^*(\neg C)/P(\neg C) \neq k \quad \text{[Lemma 5]}
\]

\[
P(B|\neg C) = \frac{P(B \& \neg C)}{P(\neg C)} \quad \text{[Def of conditional probability; } P(\neg C) > 0 \text{ by regularity of the initial assignment]}
\]

\[
P^*(B|\neg C) = \frac{P^*(B \& \neg C)}{P^*(\neg C)} \quad \text{[Def of conditional probability; } P^*(\neg C) > 0 \text{ by Lemma 1]}
\]

\[
P^*(B|\neg C) = P(B|\neg C) \quad \text{[6]}
\]

\[
P^*(B \& \neg C)/P^*(\neg C) = \frac{P(B \& \neg C)/P(\neg C)}{} \quad \text{[substitution]}
\]

\[
P^*(B \& \neg C)/P(B \& \neg C) = \frac{P^*(\neg C)/P(\neg C)}{} \quad \text{[multiplying by } P^*(\neg C)/P(B \& \neg C)]
\]

\[
P^*(\neg C)/P(\neg C) = k \quad \text{[substitution from Lemma 4]}
\]

This contradicts the assumption. Hence the conditions are not simultaneously satisfiable.

Hence, all tangles are trivial. QED

**Part II: On the triviality of tangles under exclusion**

Recall that, by 4*, a tangle is trivial iff:

\[
P(C|B) = 0 \text{ (equivalently, } P(B|C) = 0) \text{ or } P(C|B) = P(B|C) = 1.
\]

The proof below shows that if the tangle is trivial in the sense that it exemplifies the exclusion condition, \( P(B|C) = 0 \), it follows that \( P(\neg B \& \neg C) = 0 \) as well – so the distinctions between \( B \) and \( \neg C \), on the one hand, and \( C \& \neg B \), on the other, collapse.

**Lemma 1:** If \( P(C|B) = 0 \), \( P(B \& C) = 0 \).

**Proof**

1) \( P(C) > 0 \) \quad [regularity of the initial distribution]

2) \( P(B|C) = \frac{P(B \& C)}{P(C)} \) \quad [def of conditional probability]

3) \( P(B \& C) = 0 \) \quad [since \( P(B|C) = 0 \)]

Note that a comparable lemma holds for \( P(B|C) = 0 \); then, also, \( P(B \& C) = 0 \).

**Theorem:** Under conditions 1-6 plus the exclusion condition, \( P(\neg B \& \neg C) = 0 \).
This figure represents the remaining probability space under the initial distribution $P$ since it has been shown by Lemma 1 that $P(B\&C) = 0$. The variables $a$, $b$, and $c$ represent $P(\neg B \& \neg C)$, $P(B)$, and $P(C)$ respectively. $P(\neg B) = a + c$; $P(\neg C) = a + b$.

In the following chart, all twelve logically possible combinations of changes in $a$, $b$, and $c$ in the shift from $P$ to $P^*$ are shown, assuming that $a > 0$. Each variable either goes up (u), goes down (d), or remains the same (s). We know from conditions 1 and 2 that $b$ and $c$ must change, so we can dispense with the third option (s) for them.

<table>
<thead>
<tr>
<th>$a$</th>
<th>$b$</th>
<th>$c$</th>
<th>Ruled out by:</th>
</tr>
</thead>
<tbody>
<tr>
<td>u</td>
<td>u</td>
<td>u</td>
<td>$a + b + c &gt; 1$</td>
</tr>
<tr>
<td>u</td>
<td>u</td>
<td>d</td>
<td>Condition 4 (rigidity of C on $\neg B$ fails)</td>
</tr>
<tr>
<td>u</td>
<td>d</td>
<td>u</td>
<td>Condition 6 (rigidity of B on $\neg C$ fails)</td>
</tr>
<tr>
<td>u</td>
<td>d</td>
<td>d</td>
<td>Condition 4 (rigidity of C on $\neg B$ fails)</td>
</tr>
<tr>
<td>d</td>
<td>u</td>
<td>u</td>
<td>Condition 6 (rigidity of B on $\neg C$ fails)</td>
</tr>
<tr>
<td>d</td>
<td>u</td>
<td>d</td>
<td>Condition 6 (rigidity of B on $\neg C$ fails)</td>
</tr>
<tr>
<td>d</td>
<td>d</td>
<td>u</td>
<td>Condition 4 (rigidity of C on $\neg B$ fails)</td>
</tr>
<tr>
<td>d</td>
<td>d</td>
<td>d</td>
<td>$a + b + c &lt; 1$</td>
</tr>
<tr>
<td>s</td>
<td>u</td>
<td>u</td>
<td>$a + b + c &gt; 1$</td>
</tr>
<tr>
<td>s</td>
<td>u</td>
<td>d</td>
<td>Condition 4 (rigidity of C on $\neg B$ fails)</td>
</tr>
<tr>
<td>s</td>
<td>d</td>
<td>u</td>
<td>Condition 4 (rigidity of C on $\neg B$ fails)</td>
</tr>
<tr>
<td>s</td>
<td>d</td>
<td>d</td>
<td>$a + b + c &lt; 1$</td>
</tr>
</tbody>
</table>

Table 1

Since every possible combination is ruled out by the conditions given, the assumption that $a > 0$
is false. But \( a \) represents a probability, namely \( P(\neg B & \neg C) \), and therefore it cannot be negative. Therefore, \( a = 0 \). So \( P(\neg B & \neg C) = 0 \). QED.

Part III: Triviality explained

Consider the triviality condition under which \( P(B|C) = P(C|B) = 1 \). Under this condition, \( P(C & \neg B) = 0 \) and \( P(\neg C & B) = 0 \). Therefore, \( \{B & C, \neg B & \neg C\} \) is a partition of the probability space.

Therefore, this partition can be treated as a node of the evidence tree in itself. Since \( B \) and \( C \) are coextensive and \( \neg B \) and \( \neg C \) are coextensive, this node can be treated as a single variable – either simply as \( B \), \( \neg B \) or simply as \( C \), \( \neg C \). The nodes for \( B \) and for \( C \) have collapsed into a single node. Thus, evidence \( A \) that influences both \( B \) and \( C \), where the relevant screening and relevance conditions hold to create a “tangle,” influences both together, and a line in the evidence tree can be drawn from evidence \( A \) to this node alone. There is no need to try to draw arrows from \( B \) to \( C \) or vice versa to show the impact of \( A \) on both, and such arrows would have no meaning if one did draw them, as \( A \) is not influencing one by way of the other. So this tangle is trivial.

A simple semantic model of this sort of triviality would be a case where two people are co-owners of the same lottery ticket and where neither owns any other ticket. Hence, either both win or both lose, and the node can be thought of simply in terms of the winning or losing of either person (or the winning or losing of the ticket they both own).

Consider the other triviality condition under which \( P(B|C) = P(C|B) = 0 \). Under this condition, \( P(B & C) = 0 \) and (even more interestingly), modulo the “tangle” conditions, \( P(\neg B & \neg C) = 0 \). Therefore, \( \{B & \neg C, C & \neg B\} \) is a partition of the probability space.

Therefore, this partition also can be treated as a single node of the evidence tree. And, since \( B \) and \( \neg C \) are coextensive and \( C \) and \( \neg B \) are coextensive, this node can be treated as a single variable, as above. Here, too, though in a different way, the nodes for \( B \) and for \( C \) have collapsed into a single node.

A simple example of this sort of triviality would be a case of a lottery with a guaranteed winner having only two people entered, each with a different ticket. They cannot both win and they cannot both lose. So, again, the node can be thought of in terms simply of the winning or losing of one ticket-holder.
References


1. Rigidity requires that the conditional probabilities—in this case $P_{\text{old}}(H_2|H_1)$ and $P_{\text{old}}(H_2|\neg H_1)$—must be the same in the new distribution. When this condition is satisfied, the Jeffrey formula is simply a special case of the Theorem on Total Probability. It has been notoriously hard to specify conditions in which the rigidity conditions are satisfied, but this is because of Jeffrey's own desire to have the change from the old to the new distribution arise as a probabilistic surd from experience without the addition of new certain evidence. When the shift is induced by the acquisition of a new certainty, as in the cases where we wish to use it, screening of the new certain evidence by both the assertion and the negation of $H_1$ guarantees the relevant rigidity conditions.

2. It has been pointed out to us by a reviewer for Erkenntnis that there is a parallel to what we are doing (when screening conditions are fulfilled) in a formula for what is known as "convolution."

3. More generally, we should call these “lines of evidence,” since evidence can be either positively or negatively relevant.

4. We discuss this issue at greater length in our forthcoming book (identifying reference omitted).

5. This condition blocks the swelling of a basis by the addition of irrelevant sub-components. If, in the context, $C$ is quite irrelevant to $H$, then there is no point in tacking $C$ or its negation onto the end of the propositions that constitute the basis, even if the resulting partition would also screen $H$ from the given evidence.


7. Hawthorne uses $e$ to stand for an observation or experience, which, thus described, need not be propositional in nature. Our own position is that foundational evidence should be thought of as propositional and certain and as justified by experience in virtue of the fact that we have direct referential access to our own experiences (identifying reference omitted).

8. It will, of course, be pertinent to $P(H_2)$, but only in the sense that we can calculate $P(H_2)$ in terms of $P(H_1)$ and the relevant conditional probabilities. But the question of how $H_1$ supports $H_2$ is different and more interesting.

9. The qualification “at time $t$” is important, though the reasons are too complex for detailed treatment here. According to the strong foundationalist model one is gaining and losing certainties across time. Therefore, a full treatment of the dynamics of belief will have to take account of the fact that probabilities of 0 and 1 are not, as in the standard Bayesian picture, immune to revision. But changes in foundational beliefs will not be driven by an updating rule.

10. This is the way that Bayesian networks operate, updating probabilities acyclically throughout a distribution from new values given at the bottom nodes. While a Bayes net is not the same thing as an evidence tree, working with Bayes nets can be useful heuristically in avoiding mistakes in building evidence trees and in seeing how evidential force is propagated. An understanding of Bayes nets can also help to clarify the fact that conditionalization can never be done twice, thus driving home the impossibility of a real and probabilistically significant "loop of support."
11. We want to distinguish the notion of basing from the notion of confirmation. This is important because we want to discuss negative as well as positive relevance. But there is a more complicated reason for maintaining the distinction. Tomoji Shogenji (2003) has proven that, when $\pm Y$ screens off $X$ from $Z$, and the inequalities $P(Y|X) > P(Y)$ and $P(Z|Y) > P(Z)$ both hold, then $P(Z|X) > P(Z)$. In the case where a given partition acts as a basis for $e$ with respect to $H$, the items in the partition that take the place of $\pm Y$ might be complex, and although there are several useful extensions of Shogenji’s theorem for such cases, they all require some additional constraints for the transitivity of confirmation to hold with regard to any particular member of the partition.

12. We owe this objection to an anonymous reviewer for *Erkenntnis*.

13. This is almost certainly a case where a detailed analysis of the situation requires a complex node. We cannot say that the negation of "4 across is 'ruby'" screens the impact of the clue from the proposition "2 down is 'irate'," as 4 across might be, for example, 'opal' or 'jade'. The clue raises the probability that 4 across is one of those words (with 'ruby' being the best fit for the clue), and if we were given that 4 across is not 'ruby', we would still have to consider the negative impact of 'opal' or 'jade' upon the proposition that 2 down is 'irate'. Hence, it appears that what screens the evidence of the clue from the claim that 2 down is 'irate' will have to be a partition consisting of, e.g., "4 across is 'ruby' (hence not 'jade' or 'opal')," "4 across is 'jade' (hence not 'ruby' or 'opal'),," and so forth. In other words, the node in question will have to take account of all the four-letter words from the domain of possibilities that are compatible with the clue. But this technical point does not change the fact that a node consisting of statements about the specific word represented by 4 across channels the independent impact of the clue indirectly to various possibilities for 2 down.


15. This set of conditions--screening with relevance--is known as "strong screening." The example provides a counterexample to the conjecture that strong screening is transitive. For example, though $\pm Y$ strongly screens $X$ from $Z$ and $\pm Z$ strongly screens $Y$ from $W$, it is not the case that $\pm Z$ strongly screens $X$ from $W$; in fact $\pm Z$ does not screen $X$ from $W$.

16. The posteriors $P(W|\pm Z)$ are not rigid from the old distribution (before $F$ was given) to the new one. See note 1.

17. It is tempting to think that there must be some function on $Y$, $W$ such that a node could be drawn through a partition on that function – e.g. $(Y \& W)$ – rather than through $Y$ and $W$ separately. This, however, is not the case. In fact, *none* of the possible combinations $(Y \& W)$, $(Y \& \neg W)$, etc., screens $X$ from $Z$. Modulo any of these combinations, $X$ affects $Z$ directly. This curious fact is semantically explicable, but the explanation would be tedious to spell out.