

Belief Expansion, Contextual Fit and the Reliability of Information Sources

Luc Bovens and Stephan Hartmann*

Abstract

We develop a probabilistic criterion for belief expansion that is sensitive to the degree of contextual fit of the new information to our belief set as well as to the reliability of our information source. We contrast our approach with the success postulate in AGM-style belief revision and show how the idealizations in our approach can be relaxed by invoking Bayesian-Network models.

*A slightly revised version of this article appeared in V. Akman et al. (eds.), *Modeling and Using Context*, Berlin: Springer 2001, 421–424.

1 A Parable about Belief Expansion

Suppose that we have acquired various items of information from various sources and that our degree of confidence in the content of the information set is sufficiently high to believe the information. Now a new item of information is being presented by a new information source. Are we justified to add this new item of information to what we already believe? Consider the following parable:

“I go to a lecture about wildlife in Greenland which was supposed to be delivered by an expert in the field. When I arrive, I notice that the expert has excused himself and that the biology department has sent a newcomer to fill in for him. I have no beliefs about wildlife in Greenland, but I do have some beliefs about Greenland’s climate and about the kinds of climate conditions that various types of wildlife favor. Suppose that our newcomer proclaims that large colonies of elk roam in a particular valley on the southern tip of Greenland. Then I would certainly be willing to accept this item of information. But suppose that he proclaims that large colonies of armadillos roam in the same valley. Then I would not be willing to accept this item of information. Why do I accept the former but not the latter item of information? Let us suppose that neither item of information is logically inconsistent with what I already believe. Clearly, the former item of information is not inconsistent. But also the latter item is not: after all, there may be small pockets of Greenland with special climatological conditions and some Texan sailors may have set loose a pair of armadillos as a practical joke. But whereas the former item of information is quite plausible given my previous beliefs, and the latter item of information is not. And this is what makes for the difference. Now suppose that our newcomer proclaims that large colonies of wild boars roam this valley in Greenland. I am not that sure anymore. Given everything I believe, I find wild boars in Greenland more plausible than armadillos in Greenland, but certainly less plausible than elk in Greenland. Had our expert presented the lecture and provided precisely the same information, then I would have been willing to adopt the belief that there are wild boars in Greenland, but, with our newcomer delivering the lecture, I am not willing to do so. It is not that I would believe anything out of the ex-

pert's mouth: also he could not have convinced me that there are armadillos in Greenland. But when it comes to wild boars, the difference between the expert's and the newcomer's credentials simply makes for the difference."

What we learn from this story is that the question of belief expansion has something to do with the reliability of the information source as well as with my background beliefs, i.e. the context of beliefs in which this new information is supposed to be inserted. Newly presented items of information have a certain degree of contextual fit: they are more or less plausible given our background beliefs, they fit in with our background beliefs to a greater or lesser degree. The more reliable the information source is, the less contextual fit is required for me to be justified to add the belief to my belief set. The more contextual fit the new item of information has, the less reliable the information source needs to be given my context of beliefs, for me to be justified to add the belief to my belief set. The challenge is: can a precise account of this relationship be provided?

The problem is not just relevant to epistemology, but also to philosophy of science. When a scientific community is presented with new data, this does not occur within a vacuum. As is well-known from Duhem [4] and Quine [10], these new data are being assessed on the background of a context of beliefs. Whether these new data will be accepted or not is a question of the reliability of the sources (e.g. the experimental instruments) as well as of their contextual fit. Data from equally reliable sources may be accepted within one scientific community, but not in another, or in one community at some point in time, but not at another point in time, due to a difference in their degree of contextual fit. Similarly of course, data with the same degree of contextual fit may be accepted or not due to a difference in the degree of reliability of its sources. Also here an account of the precise relationship between contextual fit and the reliability of the sources is wanting.

2 The Success Postulate

The success postulate is one of the central dogmas of AGM-style belief revision: if new information comes in, then it must be incorporated into our belief set. In our parable the success postulate was clearly violated. So what

are we to make of this postulate? Certainly we could defend the postulate by pointing out that it is an idealization. In AGM-style belief revision we assume that our information sources are fully reliable and hence when new information comes in, then we are forced to incorporate it in our belief set. It is as if our information sources are of a god-like character: When God told Abraham that Sarah would bear a child, it did not seem to fit in well with what Abraham knew about Sarah's age. But since it was God who was speaking, Abraham had no choice but to incorporate it in his belief set. Similarly, if God told us that there are armadillos in Greenland, we would have no choice but to incorporate this information in our belief set. This is fair enough as an idealization. But notice that the challenge in AGM-style belief revision is how we should revise our beliefs once an inconsistency occurs in our belief set. How is it that an inconsistency can enter into our belief set when our information comes from fully reliable information sources? What remains is to phrase the idealization as follows: We assume that, although our information sources in the past may not have been fully reliable, we can rest assured that the new information that we are receiving comes from a fully reliable source. This explains why any inconsistency should be solved in favor of the incoming new information.

This idealization strikes us as rather contrived, but that is a matter of taste. More importantly, it is hard to see how a theory of belief revision that accepts the success postulate can be a truly dynamic theory that tells us how to revise our beliefs over time: Today, I must consider a new item of information as if it originated from a fully reliable information source and add it to my information set at any cost; But tomorrow, when I will be presented with a new item of information, then today's item of information will no longer be sacrosanct: any inconsistency between today's and tomorrow's item of information will need to be resolved in favor of tomorrow's item of information. Hence, tomorrow we will need to bracket the assumption that today's item of information came from a fully reliable source. But this is inconsistent: either the information came from a fully reliable source or it did not, and we cannot have it both ways. The inconsistency can be avoided if we take belief revision to model just one-shot belief revisions in the face of a single new item of information. But then it cannot be presented as a model

of how to revise our beliefs in the face of new items of information coming in over time: it would fail to be a truly dynamic theory.

The success postulate has been questioned within the belief revision literature itself. According to Hansson [6], we may not be willing to accept the new information because “it may be less reliable . . . than conflicting old information.” Makinson [7] writes that “we may not want to give top priority to new information (. . .) we may wish to weigh it against old material, and if it is really just too far-fetched or incredible, we may not wish to accept it.” Our approach is similarly motivated. However, whereas the program of non-prioritized belief revision operates within a logicist framework, we construct a probabilistic model. The cost of this approach is that it is informationally more demanding. The benefit is that it is empirically more adequate, because it is sensitive to degrees of reliability and contextual fit and to their interplay in belief acceptance. In non-prioritized belief revision, the reliability of the sources does not enter into the model itself and the lack of contextual fit of a new item of information with our belief set is understood in terms of logical inconsistency, which is only a limiting case in our model. Our model has both theoretical and practical virtues: By introducing some idealizations of our own, we reach some elegant results; At the same time, these idealizations can be readily relaxed by implementing a Bayesian Network that is responsive to the particulars of the situation. To introduce our approach, we address the simple question of belief *expansion*. We believe that our model also carries a promise to handle belief *revision* in general, but this project is beyond the scope of this paper.

3 Probabilistic Belief Expansion

Under what conditions are we justified to incorporate a new item of information in our belief set? We start with a number of idealizations about the origin of our background information - i.e. the content of my present belief set - and of our new information. The information sources are relatively unreliable, one being no more reliable than another, and independent (cf. [3]).

We stipulate that the sources are equally unreliable to simplify our model.

But why relatively unreliable? If we know that the sources are fully reliable, then what is there to discuss? Since a fully reliable source is bound to speak the truth, we have no choice but to accept and retain all the information that was provided. If we know that the sources are no better than randomizers, say, they flip coins to determine whether they will inform you that some propositions or their negations are the case, then we might as well ignore these sources altogether. So we stipulate that our new information sources are all relatively unreliable: more reliable than randomizers, but short of being fully reliable. The independence of the source is another simplification. Independent sources are sources that gather information by and only by observing the facts that they report on: they may not always provide a correct assessment of these facts, but they are not influenced by the reports of other sources, nor by the facts that other sources report on.

Let us construct a model. Suppose that there are n independent and relatively unreliable background sources and each source i informs us of a proposition R_i , for $i = 1, \dots, n$, so that the belief set is $\{R_1, \dots, R_n\}$. Furthermore, there is one independent and relatively unreliable new source informing us of a proposition R_{n+1} , so that the expanded belief set would be $\{R_1, \dots, R_{n+1}\}$. For each proposition R_i (in roman script) in the information set, let us define a propositional variable R_i (in italic script) which can take on two values, viz. R_i and \bar{R}_i (i.e. not- R_i), for $i = 1, \dots, n + 1$. Let $REPR_i$ be a propositional variable which can take on two values, viz. $REPR_i$, i.e. after consultation with the proper source, there is a report to the effect that R_i is the case, and \overline{REPR}_i , i.e. after consultation with the proper source, there is no report to the effect that R_i is the case. We construct a joint probability distribution P over $R_1, \dots, R_{n+1}, REPR_1, \dots, REPR_{n+1}$, satisfying the constraint that the sources are independent and relatively unreliable.

We model our earlier account of the independence of the sources by stipulating that P respects the following conditional independences:

$$I(\{REPR_i\}, \mathcal{M}_i | \{R_i\}) \tag{1}$$

with

$$\mathcal{M}_i = \{R_1, REPR_1, \dots, R_{i-1}, REPR_{i-1}, R_{i+1}, REPR_{i+1}, \dots, R_{n+1}, REPR_{n+1}\}$$

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

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

$$p > q > 0 \tag{2}$$

Following a tradition in epistemology that goes back to John Locke, we let belief correspond to a sufficiently high degree of confidence [5]. The degree of confidence in the background information before the new report has come in is the posterior joint probability of this information after all the relevant reports have come in:

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

Since it is assumed that we believe this information, we let $P^*(R_1, \dots, R_n)$ exceed some threshold value t for belief:

$$P^*(R_1, \dots, R_n) \geq t. \tag{4}$$

The degree of confidence in the background information conjoint with the new information is the posterior joint probability of this information after the new report has come in as well:

$$P^{**}(R_1, \dots, R_{n+1}) = P(R_1, \dots, R_{n+1} | REPR_1, \dots, REPR_{n+1}). \tag{5}$$

We are justified to expand our belief set from $\{R_1, \dots, R_n\}$ to $\{R_1, \dots, R_{n+1}\}$ just in case

$$P^{**}(R_1, \dots, R_{n+1}) \geq t. \quad (6)$$

It can be shown¹ that, for any set of propositions R_1, \dots, R_m , given the constraints on P in (1) and (2),

$$P^*(R_1, \dots, R_m) = \frac{a_0}{\sum_0^m a_i x^i}, \quad (7)$$

in which the *likelihood ratio* $x := q/p$ and a_i is the sum of the joint probabilities of all combinations of values of the variables R_1, \dots, R_n that have i negative values and $n - i$ positive values. For example, for an information triple containing the propositions R_1, R_2 , and R_3 , $a_2 = P(\bar{R}_1, \bar{R}_2, R_3) + P(\bar{R}_1, R_2, \bar{R}_3) + P(R_1, \bar{R}_2, \bar{R}_3)$. Figure 1 contains the probability space which represents a joint probability distribution over the propositional variables R_1 and R_2 and the corresponding values for a_i , for $i = 0 \dots, 2$. Note that $\sum_0^2 a_i = 1$. Suppose that the sources are twenty times as likely to report that R_i is the case, when it is the case, as then, when it is not the case, so that $x = 1/20$. Then our degree of confidence after we have received two reports is:

$$\begin{aligned} P^*(R_1, R_2) &= P(R_1, R_2 | \text{REPR}_1, \text{REPR}_2) \\ &= \frac{.15}{.15 \cdot .05^0 + .5 \cdot .05^1 + .35 \cdot .05^2} = .85 \end{aligned}$$

Now let us suppose that we receive a third item of information (i) that has a high degree of contextual fit (figure 2) and (ii) that has a low degree of contextual fit (figure 3), respectively. In case (i),

$$\begin{aligned} P'^*(R_1, R_2, R_3) &= P'(R_1, R_2, R_3 | \text{REPR}_1, \text{REPR}_2, \text{REPR}_3) \\ &= \frac{.15}{.15 \cdot .05^0 + 0 \cdot .05^1 + .5 \cdot .05^2 + .35 \cdot .05^3} = .99 \end{aligned}$$

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

whereas in case (ii),

$$\begin{aligned}
 P^{**}(R_1, R_2, R_3) &= P''(R_1, R_2, R_3 | REPR_1, REPR_2, REPR_3) \\
 &= \frac{.05}{.05 \cdot .05^0 + .3 \cdot .05^1 + .45 \cdot .05^2 + .2 \cdot .05^3} = .76
 \end{aligned}$$

Suppose that we set our threshold for acceptance at $t = .80$. Then we will have a license to expand our belief set given the likelihood ratio x that characterizes the reliability of our sources in case (i), but not in case (ii). However, we increase the reliability of the sources, i.e. if we lower the value of x , then we can boost $P^{**}(R_1, R_2, R_3)$ above the threshold level, and will have a license to expand our belief set in both cases.²

4 Bayesian Networks

The theory of Bayesian Networks is a theory of probabilistic reasoning in artificial intelligence (e.g. [8]). 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. *Root nodes* are unparented nodes and *descendant nodes* are child nodes, or child nodes of child nodes etc. There is a certain heuristic that governs the construction of the graph: there is an arrow between two nodes iff the variable in the parent node has a *direct influence* on the variable in the child node. But this is only a heuristic: the arrows in the graph have a precise probabilistic meaning, which is expressed by the *Parental Markov Condition*:

²In this paper we assess the relation of contextual fit that holds between a proposition and the belief set of a doxastic agent. In [1, 2], we develop a complementary line: we define a probabilistic measure that assesses the property of the internal coherence of a set of propositions.

(PMC) Each variable in a child node in the network is independent of all the variables in its non-descendant nodes in the network, conditional on the variables in its parent nodes. ([9], p. 19)

A Bayesian Network contains a probability distribution for the variable in each root 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. It is a central theorem in the theory of Bayesian Networks that this information is sufficient to (re)construct the joint probability distribution over all the variables in the network. Considering that the joint probability distributions for n binary variable contains 2^n entries, it is easy to see that a Bayesian Network is an economical representation of the joint probability distribution. When implemented on a computer, a Bayesian Network performs complex probabilistic calculations with one key stroke.

Let us first focus on the fact-variables R_1, \dots, R_{n+1} in figure 4. We represent the joint probability distribution over the fact-variables by means of the gray subgraph so that this DAG respects the parental Markov condition and we add a probability distribution for the variables in the root nodes and a conditional probability distribution for the variables in the child nodes.

The nodes in dotted print represent the report variables $REPR_1, \dots, REPR_{n+1}$. Each report is directly influenced by and only by the fact variables that it reports on: we draw an arrow from each fact-variable R_i to its corresponding report-variable $REPR_i$. Considering the conditional independences in (1), these arrows respect the parental Markov condition, since each child variable $REPR_i$ is indeed independent of all its non-descendants $R_1, REPR_1, \dots, R_{i-1}, REPR_{i-1}, R_{i+1}, REPR_{i+1}, \dots, R_{n+1}, REPR_{n+1}$, conditional on the parent variable R_i . We introduce the following conditional probability distribution for the variables $REPR_i$ in the dotted child nodes: $P(REPR_i|R_i) = p$ and $P(REPR_i|\bar{R}_i) = q$.

Finally, we add the white child node with the variable $R_1 \& \dots \& R_n$ which is parented by the nodes with the variables R_1, \dots, R_n . The conditional probability of $R_1 \& \dots \& R_n$ equals 1 given that R_1, \dots, R_n are the case and equals 0 otherwise (e.g. given that and). The white child node with the variable $R_1 \& \dots \& R_n$ is parented by the nodes $R_1 \& \dots \& R_n$ and R_{n+1} .

The conditional probability of $R_1 \& \dots \& R_n$ equals 1 given that R_1, \dots, R_n and R_{n+1} are the case and equals 0 otherwise.

This simple Bayesian Network permits us to read off the posterior probability of $R_1 \& \dots \& R_n$: we instantiate $REPR_1, \dots, REPR_n$, let this evidence propagate through the network and read off the posterior probability of $R_1 \& \dots \& R_n$. Subsequently, we instantiate $REPR_{n+1}$, let the evidence propagate through the network and read off the posterior probability of $R_1 \& \dots \& R_{n+1}$. We assume that the posterior probability of $R_1 \& \dots \& R_n$ exceeds the threshold value, since R_1, \dots, R_n are contained in our belief set. We may expand our belief set with R_{n+1} just in case the posterior probability of $R_1 \& \dots \& R_{n+1}$ exceeds the threshold value as well.

The advantage of bringing in Bayesian Networks is that it readily permits us to relax the idealizations. Suppose that there are more and less reliable sources in play: it is easy to adjust the $P(REPR_i|R_i)$ and $P(REPR_i|\bar{R}_i)$ for particular values of i . Suppose that two or more propositions came from the same source: we can let one source-variable parent more than one fact-variable. Suppose that the sources are not fully independent: If a source is influenced in its report by other facts than the one that it is meant to report on, we can add arrows from these other fact variables to the source variable in question and define the appropriate conditional probabilities; If a source is influenced by what other sources have to report, we can add arrows from these source variables to the source variables in question and define the appropriate conditional probabilities.

5 Conclusion

In AGM-style belief revision, the success postulate states that new information must be integrated in our belief set. This postulate is at best an extremely strong idealization. In reality, the following is the case: whether we integrate new information in our belief set is typically determined by the degree of contextual fit of this new information, i.e. by how plausible this new item of information, is given what we already believe, and by our assessment of the degree of reliability of the information sources. Similarly, in philosophy of science, the question whether new data are accepted or not

within a scientific community is a function of its contextual fit as well as of the reliability of the sources for these data. We construct a probabilistic model for belief expansion that incorporates both types of considerations. Although our model is also subject to certain idealizations, we show how these idealizations can readily be relaxed by invoking Bayesian Networks.

References

- [1] Bovens, L., Hartmann, S.: Coherence Belief Expansion and Bayesian Networks. Proceedings of the 8th International Workshop on Non-Monotonic Reasoning, NMR'2000 Breckenridge, Colorado, USA, April 9-11, 2000. (<http://www.cs.engr.uky.edu/nmr2000/proceedings.html>) (2000)
- [2] Bovens, L., Hartmann, S.: Solving the Riddle of Coherence. *Mind* **112**, 601–634 (2003).
- [3] Bovens, L., Olsson, E.: Coherentism, Reliability and Bayesian Networks. *Mind* **109**, 685–719 (2000).
- [4] Duhem, P.: The Aim and Structure of Physical Theory. trans. from the French by Philip P. Wiener (Princeton, N.J.: Princeton University Press) (1954 [1906]).
- [5] Foley, R.: The Epistemology of Belief and the Epistemology of Degrees of Belief. *American Philosophical Quarterly* **29** 111–121 (1992).
- [6] Hansson, S.O. (1997) Editor's Introduction: What's New Isn't Always Best. *Theoria* **63**, 1–13 (1997).
- [7] Makinson, D. (1997) Screened Revision. *Theoria* **63**, 14-23 (1997).
- [8] Pearl, J.: Probabilistic Reasoning in Intelligent Systems. San Mateo, CA: Morgan Kaufmann 1988.
- [9] Pearl, J.: Causality. Cambridge University Press, New York 2000.
- [10] Quine, W.V.: Two Dogmas of Empiricism. *Philosophical Review* **60**, 20-34 (1951).