Competition for Evidential Support

Gilbert Harman

Department of Philosophy, Princeton University

Abstract

In order to accept a hypothesis on the grounds that it is the best explanation of the evidence, one must know what other hypotheses compete for evidential support with the first hypothesis. But the principles for determining when hypotheses compete are obscure and represent a currently unsolved problem for this form of inference. Competing hypotheses need not contradict each other. A defender of inference to the best explanation as a distinctive form of inference will not want to identify competing hypotheses with hypotheses that are jointly highly improbable. Relying on probabilities to solve this problem would be to put the cart before the horse, since the idea behind taking inference to the best explanation to be a distinctive form of inference is that we use inference to the best explanation to determine probabilities, not the reverse Furthermore, it does not work to rely on a failure by the "hypothesis generator" to generate anything but competing hypotheses, because that just pushes the problem over to the hypothesis generator. Anyway, noncompeting hypotheses often have to be considered and therefore have to be generated Whether there is competition between two hypotheses seems to depend at least in part with whether one might be used to "fill out" the other without leading to a major change in the explanation. But it remains unclear how to distinguish "filling out" an explanation from changing it.

Keywords: explanation, inference, hypothesis, competition

The problem

Inference to the best explanation occurs when one infers the best of competing explanations of the evidence. This raises several issues. One concerns the criteria by which one out of several competing explanations is selected as the "best" of that group. Various suggestions might be made about this (Thagard, 1978, forthcoming; Harman, 1986, Harman et al., 1987, Harman et al., 1988) and I won't try to say anything about it here. Instead, I want to call attention to a different issue, namely, what makes it true that certain possible explanations are in competition for the support of that evidence?

Not all possible explanations compete in this sense Suppose Dan the detective is investigating Albert's cause of death. The evidence consists of various features of the body, the color of the skin, bruises here and there, discolorations, and also certain aspects of the surroundings, broken glass, a piece of string, and so forth Consider the following hypotheses:

(1) Albert died because he was strangled
(2) Albert died because he was poisoned

(3) Albert died because his heart stopped beating
(4) Albert died because of lack of oxygen going to his brain

Normally, under these circumstances, Dan would suppose that the strangling hypothesis competes with the poisoning hypothesis. In other words, the evidence will normally support the poisoning hypothesis only if that hypothesis provides a better explanation of the evidence (in accordance with whatever the relevant criteria are) than the strangling hypothesis does.

But normally the poisoning hypothesis will *not* compete with the heart stopping hypothesis or the lack of oxygen hypothesis. It is just not true that Dan can accept the poisoning hypothesis only if it provides a better explanation of the evidence than the heart stopping hypothesis. Dan may not know just how the poisoning might work, but in the absence of further information Dan might suppose that the heart stopping hypothesis may very well be quite compatible with the poisoning hypothesis and indeed that both might be part of a fuller explanation of the death. Therefore, Dan can envision accepting the poisoning hypothesis as the best explanation of the evidence without supposing that the poisoning hypothesis offers a better explanation than the heart stopping hypothesis.

In general, some possible explanations of the data compete in the relevant sense and some do not. The question is how to determine for any given hypothesis what its competitors are with respect to specified evidence. Such a determination is necessary before a hypothesis can be accepted as providing the best of competing explanations of the evidence.

The question of how to determine competitors raises related problems in psychology, in artificial intelligence, and in philosophy. For psychology, there is the question, "What leads people to treat possible explanations as competing?" In an artificial intelligence system using inference to the best explanation, there is the question, "How is it to be determined what to take as competing hypotheses?" For philosophy, the problem concerns which hypotheses one is "justified" in taking to be competitors for purposes of inference to be best explanation. In the philosophy of science, one question concerns what scientists actually do about this and a related question concerns what they "ought" to do about it.

In part the question concerns what hypotheses should be "generated" in considering what inference to make. But the question also arises when someone else suggests a further possible explanation that one has not or would not normally generate oneself. Sometimes one accepts (or "ought" to accept) the further suggestion as a competitor, sometimes not.

Competition as Contradiction

One natural suggestion might be that hypotheses compete if they contradict each other, given background assumptions and the evidence. This suggestion would account

for many cases For example, Kepler's hypothesis that the planets revolve around the sun in elliptical orbits competes with the hypothesis that the planets revolve around the sun in circular orbits. Here the competing hypotheses contradict each other, on the assumption that a circle does not count as an ellipse with coinciding foci.

But in many other cases competing hypotheses do not have to contradict each other. One possible hypothesis about Albert's death is that he was strangled. Another is that he was poisoned. These hypotheses compete for support from Dan's evidence, but they are not inconsistent, since Albert may have been both strangled and poisoned.

It might be objected that the hypotheses are not just (a) that Albert was poisoned and (b) that he was strangled but (a) that he died because he was poisoned and (b) that he died because he was strangled. But (a) and (b) don't contradict each other. Both might be correct. It might be that Albert's death was the overdetermined result of both causes: Sam poisoned Albert and then, just to make sure, Sam strangled him too. Either the poisoning or the strangling would have led to death at exactly the same moment. The death is equally the result of both causes! Alternatively, it could have happened that Sam's poisoning of Albert caused the death by causing someone else to strangle Albert, or vice versa! There are many possibilities of this sort, enough to show that the hypotheses of death by poisoning and death by strangulation are not literally inconsistent given the evidence and background information.

Green and grue

If the evidence is that all examined emeralds are green, one possible hypothesis is that this is so because all emeralds are green. A competing hypothesis is that all emeralds are "grue," that is, green if examined before the year 2000 and otherwise blue (Goodman, 1965). But the hypothesis that all emeralds are green does not contradict the hypothesis that all emeralds are grue, since both hypotheses would be true if all emeralds were examined before the year 2000.

Goodman's solution is to say we "assume" or "posit" conflict between these hypotheses by postulating that there are emeralds that will not be examined before 2000 (Ullian & Goodman, 1976). But the hypotheses would seem to compete in the relevant sense even in the absence of such a postulate. For if we accept one of these hypotheses as the best account of the data, e.g., "All emeralds are green," then there is no need to look any farther. The data need no further explanation of the sort that might be provided by "All emeralds are grue."

Appealing to Probability

Thagard (forthcoming) notes that in the debate over dinosaur extinction, scientists take the following two hypotheses to compete: (1) Dinosaurs became extinct because of a meteorite collision. (2) Dinosaurs became extinct because the sea level fell. These

hypotheses do not contradict each other and could both be true. What is it that leads scientists to treat these hypotheses but not others as competing?

Why can't we be content with a theory of inference to the best explanation that has no principled account of when hypotheses are to be treated as competitors? Why not simply rely on scientists' judgments about when hypotheses compete? Answer: because there ought to be some account of where those judgments come from. To refuse to give a principled answer to this question is not very different from simply relying on scientific judgments about whether the evidence supports a given conclusion. If we want to give a principled account of when evidence supports a given conclusion, we need to be able to give a principled account of when hypotheses are in competition for support from particular evidence.

Thagard (forthcoming) speculates that scientists may treat (1) and (2) as competing hypotheses because (a) there are no explanatory relations between them and (b) their conjunction is unlikely. There are two parts to this suggestion, (a) and (b), which I would like to consider separately in reverse order.

Consider the suggestion that possible explanations of the evidence conflict if their conjunction is unlikely. In what way "unlikely"? If what is meant is that their conjunction is *a priori* unlikely without considering the evidence in this particular case, then almost any conjunction of hypotheses will be unlikely and almost any two hypotheses must be treated as conflicting, which is contrary to ordinary practice. On the other hand, if we are to take the evidence into account and consider the probabilities of the conjunction of the two hypotheses given that evidence, then we must ask where these probabilities come from. The thought behind the idea that there is such a thing as inference to the best explanation is that our probability judgments depend on our judgments about the relative merits of competing explanations: the standards of inference to the best explanation influence judgments of probabilities rather than the other way round. If probabilities could be determined without reference to such explanatory considerations, we would not need inference to the best explanation. But then we cannot use probabilities to decide when hypotheses compete, since we have to know what hypotheses compete in order to reach a judgment about probabilities. Thagard (forthcoming) makes a similar point in discussing Pearl's (1986, 1987) work on belief networks.

Generate Only Competitors?

One way to try to circumvent the difficulty of finding a principled account of competition in an artificial intelligence reasoning program that uses inference to the best explanation is to see to it that the program never generates anything but competitors of the hypothesis being considered. But how is this to be accomplished? The usual method is to put ad hoc restrictions on the hypotheses to be considered. But can this be done in a principled and non ad hoc way? Furthermore, what happens when a hypothesis is suggested to the system from outside, a hypothesis that is not thought up by the system

itself? How does it decide whether that hypothesis is in competition with the other hypotheses being considered.

Further consideration indicates that trying to restrict what hypotheses are generated is not going to work. Recall Detective Dan who is trying to determine the cause of Albert's death. It is quite likely that explanations like heart stopping and no oxygen to the brain will be *considered* by Dan. The hypotheses will be "generated," they just won't be generated *as* competitors of poisoning or strangling. But how are they distinguished from hypotheses that are generated as competitors of the poisoning and strangling hypotheses?

Explanatory Relations among Hypotheses

Thagard's other suggestion about competition was that scientists treat the meteorite collision hypothesis and the falling seal level hypothesis as competing hypotheses with respect to the extinction of Dinosaurs because "there are no explanatory relations between them ..." This is more promising, although there are difficulties in making the suggestion precise.

Consider again Arthur's death. The hypothesis that he died because he was poisoned does not conflict with the hypothesis that he died because his heart stopped beating The lack of conflict does not have to depend on the definite belief that there in an explanatory relation between these two hypotheses. Dan may have no idea how the poison works. He does not have to have a positive belief that poison works by stopping the heart, for example. The point is rather that he supposes that if the evidence is explained by the poisoning, it is *possible* that further explanatory work might be done by the heart stopping. He envisions that the explanation of Albert's death by poisoning might be filled in with a further account of how the poisoning led to the death, where that further explanation might involve Albert's dying because his heart stopped beating.

That is not what Dan is envisioning in the poisoning and strangling case. He does not expect the poisoning to have been caused by the strangling, or vice versa: although that is possible!

But now there seems to be a dilemma. If we say that hypotheses that might explain the evidence compete if and only if it is *not possible* for the one explanation to be filled out by using the second, then we have to say that poisoning and strangling do not compete! If we say that it is not "expected" that the one hypothesis will be filled out by using the second, then this is rarely "positively expected". If the reference to what is "expected" is a way of appealing to probability, we are back with the problem that appeal to inference to the best explanation is supposed to account of judgments of probability rather than vice versa.

Maybe the answer is this: Although Dan can imagine that someone's poisoning Albert may have caused Albert's death by leading someone else to strangle Albert, that is a

different explanation from what Dan envisions when he hypothesizes that Albert died because he was poisoned. Perhaps two hypotheses compete if the use of either to "fill in" the explanation provided by the other would yield a "different explanation" from what one is envisioning.

This raises the question of what the difference is between "filling out" an explanation, e.g. as when Dan explains how the poisoning caused Albert's death by causing Albert's heart to stop beating, and "yielding a different explanation from what was originally envisioned," e.g. as when Dan decides that the poisoning caused Albert's death by leading someone else to strangle Albert.

And how is this sort of difference something that can be used on the spot by a human being or scientist or artificial intelligence system to determine what the competitors of a given hypothesis are for the support of certain evidence the hypothesis might explain?

The preparation of this paper was supported in part by a research grant to Princeton University from the James S McDonnell Foundation.

Bibliography

Goodman, N. (1965). *Fact, Fiction, and Forecast*, 2nd edition. Indianapolis: Bobbs, Merrill.

Harman, G. (1986) *Change in View: Principles of Reasoning*. Cambridge, Massachusetts; M.I.T./Bradford Books.

Harman, G., Bienkowski, M. A., Salem, K., & Pratt, I (1987) "Measuring change and coherence in evaluating potential change in view." *Ninth Annual Conference of the Cognitive Science Society*. Hillsdale, N.J.: Erlbaum, 203-209.

Harman, G., Ranney, M., Salem, K., Döring, F., Epstein, J. & Jaworska, A. (1988). "A theory of simplicity." *Tenth Annual Conference of the Cognitive Science Society*. Hillsdale, N.J.: Erlbaum, 111-117.

Pearl, J. (1986) "Fusion, propagation, and structuring in belief networks," *Artificial Intelligence* 29: 241-288.

Pearl, J. (1987) "Distributed revision of composite beliefs," *Artificial Intelligence*, 33: 173-215.

Thagard, P. (19787). "The best explanation: criteria for theory choice." *Journal of Philosophy* 75: 76-92.

Thagard, P. (forthcoming). "Explanatory coherence." *Behavioral and Brain Sciences*.

Ullian, J. & Goodman, N. (1976). "Projectibility unscathed." *Journal of Philosophy* 73: 527-531.