Cambridge University Press
0521434610 - Abductive inference: Computation, Philosophy, Technology
Edited by John R. Josephson and Susan G. Josephson
Excerpt
More information

Introduction

I see abductive inferences everywhere in science and ordinary life. I believe abductions to be reasonable and knowledge-producing inferences. If this view is correct, there appear to be significant philosophical implications: It leads to a form of Realism about the objects of theory and perception; it leads to the view that Truth is attainable but extremely high certainty is not; it extends the detailed conception of Reason to better accommodate fallibility and uncertainty; it loosens the bounds on what can be known; it finds the logic of science to be very akin to reasoning in ordinary life and to the learning of children; and it moves toward restoring confidence in objectivity and progress where it has been most deeply threatened.

I have been thinking about abduction since the mid-1970s, when my doctoral philosophy of science dissertation project on Causality transmuted itself into a sustained attempt to reconstruct the logical foundations of science based on Gilbert Harman's idea of inference to the best explanation. I argued that abductive foundations are stronger than those based on induction, and that there are conceptual advantages to this view for a number of traditional philosophical puzzles, including the problem of induction.

My dissertation found abductive inferences in ordinary life as well as at the foundations of science and argued that they are epistemically warranted. It developed a process view of inference, rather than a static, evidential-relationship view, but it did not yet take a computational view. Although I discussed a research program based on trying to build robot scientists, I had not yet begun this kind of work.

I finished the dissertation in 1982 and promptly began learning Artificial Intelligence (AI) with B. Chandrasekaran (Chandra) at The Ohio State University. From Chandra I learned to take a computational, or better, an "information-processing," view of knowledge and intelligence. His research program, which I embraced, was engaged in a search for fundamental building blocks of intelligence that were expected to take the form of "generic information-processing tasks."

My dissertation proposed investigating the inferential practices of science by trying to design robot scientists, and seeing what it would take to

By John R. Josephson.

make them work. Still, at the time it was written, I wasn't sensitive to the implications of taking an information-processing view of inference and intelligence. Since then Chandra's tutelage in AI and my experiences in designing and building knowledge-based systems have significantly enriched my view. They have especially sensitized me to the need to provide for feasible computation in the face of bounded computational resources. To provide for feasible computation, a model of intelligence must provide for the control of reasoning processes, and for the organization and representation of knowledge.

When I joined the AI group at Ohio State (which later became the Laboratory for Artificial Intelligence Research, or LAIR), it was intensely studying diagnosis and looking for the generic tasks hypothesized by Chandra to be the computational building blocks of intelligence. Besides Chandra and me, the AI group at the time consisted of Jack Smith, Dave Brown, Tom Bylander, Jon Sticklen, Mike Tanner, and a few others. Working primarily with medical domains, the group had identified "hierarchical classification" as a central task of diagnosis, had distinguished this sort of reasoning from "data abstraction" and other types of reasoning that enter into diagnostic problem solving, and was trying to push the limits of this view by attempting new knowledge domains.

My first major project with the AI group in 1983 was a collaboration with Jack Smith, MD, and others, on the design and construction of a knowledge-based system (called RED) for an antibody-identification task performed repeatedly by humans in hospital blood banks. The task requires the formation of a composite mini-theory for each particular case that describes the red-cell antibodies present in a patient's blood. Our goal was to study the problem-solving activity of an expert and to capture in a computer program enough of the expert's knowledge and reasoning strategy to achieve good performance on test cases. Preenumerating all possible antibody combinations would have been possible (barely) but this was forbidden because such a solution would not scale up.

The reasoning processes that we were trying to capture turned out to include a form of best-explanation reasoning. Before long it became clear that classification was not enough to do justice to the problem, that some way of controlling the formation of multipart hypotheses was needed. This led me to design what we now call the RED-1 hypothesis-assembly algorithm. We then built RED-1, a successful working system with a novel architecture for hypothesis formation and criticism. RED-1's successor, RED-2, was widely demonstrated and was described in a number of papers. Jack Smith (already an MD) wrote his doctoral dissertation in computer science on the RED work.

The RED systems show that abduction can indeed be made precise enough to be a usable notion and, in fact, precise enough to be programmed. These systems work well and give objectively good answers, even in complicated cases where the evidence is ambiguous. They do not manipulate numerical probabilities, follow deductive inference rules, or generalize from experience. This strongly reinforces the argument that abduction is a distinct form of inference, interesting in its own right.

RED-1 was the first, and RED-2 the second, of six generations of abductive-assembly mechanisms that we designed and with which we experimented. In the following chapters the evolution of these machines is traced as they grew in power and sophistication. They were all intended as domain-independent abductive problem solvers, embodying inference-control strategies with some pretensions of generality. One design for parallel hypothesis assembly was never implemented, but each of the other five mechanisms was implemented, at least partially, and a fair amount of experience was built up in our lab with abductive problem solving.

In the PEIRCE project (named after Charles Sanders Peirce) we made generalizations and improvements to RED-2's hypothesis-assembly mechanism. PEIRCE is a domain-independent software tool for building knowledge-based systems that form composite explanatory hypotheses as part of the problem-solving process. PEIRCE has various hypothesis-improvement tactics built in and allows the knowledge-system builder to specify strategies for mixing these tactics. Members of our group also designed and built other abductive systems, including MDX2 by Jon Sticklen, TIPS (Task Integrated Problem Solver) by Bill Punch, and QUAWDS (Qualitative Analysis of Walking Disorders) by Tom Bylander and Mike Weintraub. Other discoveries were made in collaboration with Mike Tanner, Dean Allemang, Ashok Goel, Todd Johnson, Olivier Fischer, Matt DeJongh, Richard Fox, Susan Korda, and Irene Ku. Most of the abduction work has been for diagnosis in medical and mechanical domains, but more recently, in collaboration with several speech scientists and linguists here at Ohio State, we have begun to work on layered-abduction models of speech recognition and understanding.

Susan Josephson has been my mate and a stimulating intellectual companion throughout my adult life. When the project of editing this book bogged down in the summer of 1990, Susan agreed to take the lead and set aside for a time her project of writing a book on the philosophy of AI. The present book is a result of our collaboration and consists of a deeply edited collection of LAIR writings on abduction by various authors. I take responsibility for the major editorial decisions, especially the controversial ones. Susan is responsible for transforming a scattered set of material into a unified narrative and sustained argument, and she produced the first draft. Many voices blend in the text that follows, although mine is the most common. Authors whose work is included here should not be presumed to agree with all conclusions.

In chapter 1 we set the stage with a careful discussion of abduction and some of its relationships with other traditionally recognized forms of inference. This is followed in chapter 2 by an orientation to our view of AI as a science and to our approach to building knowledge systems. The remainder of the book traces the development of six generations of abduction machines and describes some of the discoveries that we made about the dynamic logic of abduction.

1 Conceptual analysis of abduction

What is abduction?

Abduction, or *inference to the best explanation*, is a form of inference that goes from data describing something to a hypothesis that best explains or accounts for the data. Thus abduction is a kind of theory-forming or interpretive inference. The philosopher and logician Charles Sanders Peirce (1839–1914) contended that there occurs in science and in everyday life a distinctive pattern of reasoning wherein explanatory hypotheses are formed and accepted. He called this kind of reasoning "abduction."

In their popular textbook on artificial intelligence (AI), Charniak and McDermott (1985) characterize abduction variously as modus ponens turned backward, inferring the cause of something, generation of explanations for what we see around us, and inference to the best explanation. They write that medical diagnosis, story understanding, vision, and understanding natural language are all abductive processes. Philosophers have written of "inference to the best explanation" (Harman, 1965) and "the explanatory inference" (Lycan, 1988). Psychologists have found "explanation-based" evidence evaluation in the decision-making processes of juries in law courts (Pennington & Hastie, 1988).

We take abduction to be a distinctive kind of inference that follows this pattern pretty nearly:[1]

> *D* is a collection of data (facts, observations, givens).
> *H* explains *D* (would, if true, explain *D*).
> No other hypothesis can explain *D* as well as *H* does.
>
> Therefore, *H* is probably true.

The core idea is that a body of data provides evidence for a hypothesis that satisfactorily explains or accounts for that data (or at least it provides evidence if the hypothesis is better than explanatory alternatives).

Abductions appear everywhere in the un-self-conscious reasonings, inter-

This chapter was written by John R. Josephson, except the second section on diagnosis, which was written by Michael C. Tanner and John R. Josephson.

pretations, and perceivings of ordinary life and in the more critically self-aware reasonings upon which scientific theories are based. Sometimes abductions are deliberate, such as when the physician, or the mechanic, or the scientist, or the detective forms hypotheses explicitly and evaluates them to find the best explanation. Sometimes abductions are more perceptual, such as when we separate foreground from background planes in a scene, thereby making sense of the disparities between the images formed from the two eyes, or when we understand the meaning of a sentence and thereby explain the presence and order of the words.

Abduction in ordinary life

Abductive reasoning is quite ordinary and commonsensical. For example, as Harman (1965) pointed out, when we infer from a person's behavior to some fact about her mental state, we are inferring that the fact explains the behavior better than some other competing explanation does. Consider this specimen of ordinary reasoning:

JOE: Why are you pulling into the filling station?
TIDMARSH: Because the gas tank is nearly empty.
JOE: What makes you think so?
TIDMARSH: Because the gas gauge indicates nearly empty. Also, I have no reason to think that the gauge is broken, and it has been a long time since I filled the tank.

Under the circumstances, the nearly empty gas tank is the best available explanation for the gauge indication. Tidmarsh's other remarks can be understood as being directed to ruling out a possible competing explanation (broken gauge) and supporting the plausibility of the preferred explanation.

Consider another example of abductive reasoning: Imagine that one day you are driving your car, and you notice the car behind you because of its peculiar shade of bright yellow. You make two turns along your accustomed path homeward and then notice that the yellow car is still behind you, but now it is a little farther away. Suddenly, you remember something that you left at the office and decide to turn around and go back for it. You execute several complicated maneuvers to reverse your direction and return to the office. A few minutes later you notice the same yellow car behind you. You conceive the hypothesis that you are being followed, but you cannot imagine any reason why this should be so that seems to have any significant degree of likelihood. So, you again reverse direction, and observe that the yellow car is still behind you. You conclude that you are indeed being followed (reasons unknown) by the person in the dark glasses in the yellow car. There is no other plausible way to explain why the car remains continually behind you. The results of your experiment of reversing direction a second time served to rule out alternative explanations, such as that the other driver's first reversal of direction was a coincidence of changing plans at the same time.

Harman (1965) gave a strikingly insightful analysis of law court testimony, which argues that when we infer that a witness is telling the truth, we are using best-explanation reasoning. According to Harman our inference goes as follows:

(i) We infer that he says what he does because he believes it.

(ii) We infer that he believes what he does because he actually did witness the situation which he describes.

Our confidence in the testimony is based on our conclusions about the most plausible explanation for that testimony. Our confidence fails if we come to think that there is some other plausible explanation for his testimony – for example, that he stands to gain from our believing him. Here, too, we see the same pattern of reasoning from observations to a hypothesis that explains those observations – not simply to a possible explanation, but to the best explanation for the observations in contrast with alternatives.

In *Winnie-the-Pooh* (Milne, 1926) Pooh says:

> It had HUNNY written on it, but, just to make sure, he took off the paper cover and looked at it, and it *looked* just like honey. "But you never can tell," said Pooh. "I remember my uncle saying once that he had seen cheese just this colour." So he put his tongue in, and took a large lick. (pp. 61–62)

Pooh's hypothesis is that the substance in the jar is honey, and he has two pieces of evidence to substantiate his hypothesis: It looks like honey, and "hunny" is written on the jar. How can this be explained except by supposing that the substance is honey? He considers an alternative hypothesis: It might be cheese. Cheese has been observed to have this color, so the cheese hypothesis offers another explanation for the color of the substance in the jar. So, Pooh (conveniently dismissing the evidence of the label) actively seeks evidence that would distinguish between the hypotheses. He performs a test, a crucial experiment. He takes a sample.

The characteristic reasoning processes of fictional detectives have also been characterized as abduction (Sebeok & Umiker-Sebeok, 1983). To use another example from Harman (1965), when a detective puts the evidence together and decides that the culprit *must* have been the butler, the detective is reasoning that no other explanation that accounts for all the facts is plausible enough or simple enough to be accepted. Truzzi (1983) alleges that at least 217 abductions can be found in the Sherlock Holmes canon.

> "There is no great mystery in this matter," he said, taking the cup of tea which I had poured out for him; "the facts appear to admit of only one explanation."
>
> – Sherlock Holmes (Doyle, 1890, p. 620)

Abduction in science

Abductions are common in scientific reasoning on large and small scales.[2] The persuasiveness of Newton's theory of gravitation was enhanced by its ability to explain not only the motion of the planets, but also the occurrence of the tides. In *On the Origin of Species by Means of Natural Selection* Darwin presented what amounts to an extended argument for natural selection as the best hypothesis for explaining the biological and fossil evidence at hand. Harman (1965) again: when a scientist infers the existence of atoms and subatomic particles, she is inferring the truth of an explanation for her various data. *Science News* (Peterson, 1990) reported the attempts of astronomers to explain a spectacular burst of X rays from the globular cluster M15 on the edge of the Milky Way. In this case the inability of the scientists to come up with a satisfactory explanation cast doubt on how well astronomers understand what happens when a neutron star accretes matter from an orbiting companion star. *Science News* (Monastersky, 1990) reported attempts to explain certain irregular blocks of black rock containing fossilized plant matter. The best explanation appears to be that they are dinosaur feces.

Abduction and history

Knowledge of the historical past also rests on abductions. Peirce (quoted in Fann, 1970) cites one example:

> Numberless documents refer to a conqueror called Napoleon Bonaparte. Though we have not seen the man, yet we cannot explain what we have seen, namely, all those documents and monuments without supposing that he really existed. (p. 21)

Abduction and language

Language understanding is another process of forming and accepting explanatory hypotheses. Consider the written sentence, “The man sew the rat eating the corn.” The conclusion seems inescapable that there has been some sort of mistake in the third word “sew” and that somehow the “e” has improperly replaced an “a.” If we are poor at spelling, or if we read the sentence rapidly, we may leap to the “saw” reading without even noticing that we have not dealt with the fact of the “e.” Taking the “saw” reading demands our acceptance so strongly that it can cause us to overturn the direct evidence of the letters on the page, and to append a hypothesis of a mistake, rather than accept the hypothesis of a nonsense sentence.

The process of abduction

Sometimes a distinction has been made between an initial process of coming up with explanatorily useful hypothesis alternatives and a subsequent process of critical evaluation wherein a decision is made as to which explanation is best. Sometimes the term “abduction” has been restricted to the hypothesis-generation phase. In this book, we use the term for the whole process of generation, criticism, and acceptance of explanatory hypotheses.

One reason is that although the explanatory hypotheses in abduction can be simple, more typically they are composite, multipart hypotheses. A scientific theory is typically a composite with many separate parts holding together in various ways,[3] and so is our understanding of a sentence and our judgment of a law case. However, no feasible information-processing strategy can afford to explicitly consider all possible combinations of potentially usable theory parts, since the number of combinations grows exponentially with the number of parts available (see chapter 7). Reasonably sized problems would take cosmological amounts of time. So, one must typically adopt a strategy that avoids generating all possible explainers. Prescreening theory fragments to remove those that are implausible under the circumstances makes it possible to radically restrict the potential combinations that can be generated, and thus goes a long way towards taming the combinatorial explosion. However, because such a strategy mixes critical evaluation into the hypothesis-generation process, this strategy does not allow a clear separation between the process of coming up with explanatory hypotheses and the process of acceptance. Thus, computationally, it seems best not to neatly separate generation and acceptance. We take *abduction* to include the whole process of generation, criticism, and possible acceptance of explanatory hypotheses.

Diagnosis and abductive justification

In this section we show by example how the abductive inference pattern can be used simply and directly to describe diagnostic reasoning and its justifications.

In AI, diagnosis is often described as an abduction problem (e.g., Peng & Reggia, 1990). Diagnosis can be viewed as producing an explanation that best accounts for the patient's (or device's) symptoms. The idea is that the task of a diagnostic reasoner is to come up with a best explanation for the symptoms, which are typically those findings for the case that show abnormal values. The explanatory hypotheses appropriate for diagnosis are malfunction hypotheses: typically disease hypotheses for plants and animals and broken-part hypotheses for mechanical systems.

The diagnostic task is to find a malfunction, or set of malfunctions, that best explains the symptoms. More specifically, a diagnostic conclusion should explain the symptoms, it should be plausible, and it should be significantly better than alternative explanations. (The terms "explain," "plausible," and "better" remain undefined for now.)

Taking diagnosis as abduction determines the classes of questions that are fair to ask of a diagnostician. It also suggests that computer-based diagnostic systems should be designed to make answering such questions straightforward.

Consider the example of liver disease diagnosis given by Harvey and Bordley (1972, pp. 299–302). In this case the physician organized the differential (the set of alternative hypotheses) around hepatomegaly (enlarged liver), giving five categories of possible causes of hepatomegaly: venous congestion of the liver, obstruction of the common duct, infection of the liver, diffuse hepatomegaly without infection, and neoplasm (tumor) of the liver. He then proceeded to describe the evidence for and against each hypothesis. Venous congestion of the liver was ruled out because none of its important symptoms were present. Obstruction of the common duct was judged to be unlikely because it would not explain certain important findings, and many expected symptoms were not present. Various liver infections were judged to be explanatorily irrelevant because certain important findings could not be explained this way. Other liver infections were ruled out because expected consequences failed to appear, although one type of infection seemed somewhat plausible. Diffuse hepatomegaly without infection was considered explanatorily irrelevant because, by itself, it would not be sufficient to explain the degree of liver enlargement. Neoplasm was considered to be plausible and would adequately explain all the important findings. Finally, the physician concluded the following:

> The real choice here seems to lie between an infection of the liver and neoplasm of the liver. It seems to me that the course of the illness is compatible with a massive hepatoma [neoplasm of the liver] and that the hepatomegaly, coupled with the biochemical findings, including the moderate degree of jaundice, are best explained by this diagnosis.

Notice the form of the argument:

1. There is a finding that must be explained (hepatomegaly).
2. The finding might be explained in a number of ways (venous congestion of the liver, obstruction of the common duct, infection of the liver, diffuse hepatomegaly without infection, and neoplasm of the liver).
3. Some of these ways are judged to be implausible because expected consequences do not appear (venous congestion of the liver).
4. Some ways are judged to be irrelevant or implausible because they do not explain important findings (obstruction of the common duct, diffuse hepatomegaly without infection).
5. Of the plausible explanations that remain (infection of the liver, neoplasm of the liver), the best (neoplasm of the liver) is the diagnostic conclusion.

The argument is an abductive justification for the diagnostic conclusion.

Suppose the conclusion turned out to be wrong. What could have happened to the true answer? That is, why was the true, or correct, answer not the best explanation? This could only have happened for one or more of the following reasons:

1. There was something wrong with the data such that it really did not need to be explained. In this case, hepatomegaly might not have actually been present.
2. The differential was not broad enough. There might be causes of hepatomegaly that were unknown to the physician, or that were overlooked by him.