

# Abductive Inference

## Computation, Philosophy, Technology

Edited by

John R. Josephson and Susan G. Josephson

© Cambridge University Press, 1996

### Chapter 1 Conceptual analysis of abduction <sup>†</sup>

<sup>†</sup> 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.

#### 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:<sup>1</sup>

*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, interpretations, 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.<sup>2</sup> 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,<sup>3</sup> 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.
3. Hypotheses were incorrectly judged to be implausible. Perhaps venous congestion should have been considered more plausible than it was, due to faulty knowledge or missing evidence.
- 4a. Hypotheses were incorrectly thought not to explain important findings. For example, obstruction might explain findings that the physician thought it could not, possibly because the physician had faulty knowledge.

- 4b. The diagnostic conclusion was incorrectly thought to explain the findings. Neoplasm might not explain the findings, due to faulty knowledge or to overlooking important findings.
- 5a. The diagnostic conclusion was incorrectly thought to be better than it was. Neoplasm might have been overrated, due to faulty knowledge or missing evidence.
- 5b. The true answer was underrated, due to faulty knowledge or missing evidence.

Many questions to the diagnostician can be seen as indicating ways in which the answer may be wrong, each question suggesting an error of a particular type. An answer to such a question should convince the questioner that the diagnosis is not mistaken in that way.

Returning to the example, if the physician were asked, “What makes venous congestion implausible?” he might answer:

This patient exhibited no evidence of circulatory congestion or obstruction of the hepatic veins or vena cava. . . .

thus trying to convince the questioner that venous congestion was correctly ruled out. If asked, “Why not consider some toxic hepatic injury?” the physician could reply:

[It would not] seem to compete with a large hepatoma in explaining the massive hepatomegaly, the hypoglycemia, and the manifestations suggestive of infection.

thus trying to convince the questioner that the differential is broad enough.

Interestingly, in this case Bordley's diagnosis *was* wrong. Autopsy revealed that the patient actually had cancer of the pancreas. (To be fair, the autopsy also found tumors in the liver, but pancreatic cancer was considered the primary illness.) One significant finding in the case was elevated amylase, which is not explained by neoplasm of the liver. So, if we asked the physician, “How do you account for the sharply elevated amylase?” his only possible reply would be:

Oops.

The diagnosis was inadequate because it failed to account for all the important findings (item 4b in the previous numbered list).

This analysis tells us that if we build an expert system and claim that it does diagnosis, we can expect it to be asked certain questions. These are the only questions that are fair to ask simply because it is a diagnostic system. Other questions would not be about diagnosis per se. These other questions might include requests for definitions of terms, exam-like questions that check

the system's knowledge about some important fact, and questions about the implications of the diagnostic conclusion for treatment. Thus, the idea of abductive justification gives rise to a model of dialogue between the diagnostician and the client. It defines a set of questions that any person, or machine, claiming to do diagnosis should be prepared to answer.

One power of this analysis lies in controlling for error, in making explicit the ways in which the conclusion can be wrong. A challenging question implies that the questioner thinks that the answer might be wrong and that the questioner needs to be convinced that it is not. A proper answer will reassure the questioner that the suspected error has not occurred.

## Doubt and certainty

### *Inference and logic*

Inferences are movements of thought within the sphere of belief.<sup>4</sup> The function of inference is the acceptance (or sometimes rejection) of propositions on the basis of purported evidence. Yet, inferences are not wholly or merely psychological; there may be objective relationships of evidential support (or its absence) between propositions that have nothing much to do with whether anyone thinks of them. Thus a science of evidential relationships is possible that has very little to do with empirical psychology. This science is *logic* in the broad sense.

### *Deduction and abduction*

Deductions support their conclusions in such a way that the conclusions must be true, given true premises; they convey conclusive evidence. Other forms of evidential support are not so strong, and though significant support for a conclusion may be given, a possibility of error remains. Abductions are of this kind; they are fallible inferences.

Consider the following logical form, commonly called *disjunctive syllogism*.

$P$  or  $Q$  or  $R$  or  $S$  or . . .

But not- $Q$ , not- $R$ , not- $S$ , . . .

---

Therefore,  $P$ .

This form is deductively valid. Moreover, the support for an abductive conclusion fits this form if we assert that we have exhaustively enumerated all possible explanations for the data and that all but one of the alternative explanations has been decisively ruled out. Typically, however, we will have

reasons to believe that we have considered all plausible explanations (i.e., those that have a significant chance of being true), but these reasons stop short of being conclusive. We may have struggled to formulate a wide variety of possible explanations but cannot be sure that we have covered all plausibles. Under these circumstances we can assert a proposition of the form of the first premise of the syllogism, but assert it only with a kind of qualified confidence. Typically, too, alternative explanations can be discounted for one reason or another but not decisively ruled out. Thus abductive inferences, in a way, rely on this particular deductively valid inference form, but abductions are conclusive only in the limit.

Of course disjunctive syllogism fits any decision by enumeration of alternatives and exclusion, not just abductions (where *explanatory* alternatives are considered). From this it can be seen that abduction cannot be identified with disjunctive syllogism.

### *Ampliative inference*

Like inductive generalizations, abductions are *ampliative inferences*; that is, at the end of an abductive process, having accepted a best explanation, we may have more information than we had before. The abduction transcends the information of its premises and generates new information that was not previously encoded there at all. This can be contrasted with deductions, which can be thought of as extracting, explicitly in their conclusions, information that was already implicitly contained in the premises. Deductions are *truth preserving*, whereas successful abductions may be said to be *truth producing*.

This ampliative reasoning is sometimes done by introducing new vocabulary in the conclusion. For example, when we abduce that the patient has hepatitis because hepatitis is the only plausible way to explain the jaundice, we have introduced into the conclusion a new term, “hepatitis,” which is from the vocabulary of diseases and not part of the vocabulary of symptoms. By introducing this term, we make conceptual connections with the typical progress of the disease, and ways to treat it, that were unavailable before. Whereas valid deductive inferences cannot contain terms in their conclusions that do not occur in their premises, abductions can “interpret” the given data in a new vocabulary. Abductions can thus make the leap from “observation language” to “theory language.”

### *Doubt and hesitation*

An abductive process aims at a satisfactory explanation, one that can be confidently accepted. However it may be accompanied in the end with some explicit qualification, for example, some explicit degree of assurance or some doubt. One main form of doubt is just hesitation from being aware of the possibility of alternative explanations. Classically, this is just how Descartes generates doubts about knowledge from the senses: Alternative explanations to

the usual interpretations of sensory information are that we are dreaming or that we are being deceived by a very powerful and evil demon (Descartes, 1641). Since low-plausibility alternative explanations can be generated indefinitely, doubt cannot be completely eliminated.

On the way to a satisfactory explanation, an abductive process might seek further information beyond that given in the data initially to be explained. For example, there may be a need to distinguish between explanatory alternatives; for help in forming hypotheses; or for help in evaluating them. Often abductive processes are not immediately concluded, but are suspended to wait for answers from information-seeking processes. Such suspensions of processing can last a very long time. Years later, someone may say, "So that's why she never told me. I was always puzzled about that." Centuries later we may say, "So that's the secret of inheritance. It's based on making copies of long molecules that encode hereditary information."

### *Abductive conclusions: likelihood and acceptance*

As we said earlier, abductions follow approximately this pattern:

$D$  is a collection of data.

$H$  explains  $D$ .

No other hypothesis can explain  $D$  as well as  $H$  does.

---

Therefore,  $H$  is probably true.

The judgment of likelihood associated with an abductive conclusion should depend on the following considerations (as it typically does in the inferences we actually make):

1. how decisively  $H$  surpasses the alternatives<sup>5</sup>
2. how good  $H$  is by itself, independently of considering the alternatives (we should be cautious about accepting a hypothesis, even if it is clearly the best one we have, if it is not sufficiently plausible in itself)
3. judgments of the reliability of the data
4. how much confidence there is that all plausible explanations have been considered (how thorough was the search for alternative explanations).<sup>6</sup>

Beyond the judgment of its likelihood, willingness to accept the conclusion should (and typically does) depend on:

1. pragmatic considerations, including the costs of being wrong and the benefits of being right

2. how strong the need is to come to a conclusion at all, especially considering the possibility of seeking further evidence before deciding.

This theory of abduction is an evaluative theory, offering standards for judging reasoning patterns. It is also a descriptive and explanatory theory for human and computer reasoning processes that provides a way of analyzing these processes in functional terms, of showing what they accomplish, of showing how they manifest good reasoning (i.e., intelligence).

*Best explanation: compared with what?*

There is some ambiguity in the abductive inference pattern as we have described it. What is the set of explanatory hypotheses of which  $H$  is the best? Perhaps this premise should say, “No other *available* hypothesis can explain  $D$  as well as  $H$  does.”<sup>7</sup> The set of alternatives might be thought of so narrowly as to include just those hypotheses that one thinks of immediately, or so broadly as to include all the hypotheses that can in principle be formulated. Construed broadly, it would include the true explanation, which would of course be best, but then the whole inference form would seem to be trivial.

Yet it appears that the force of the abductive inference depends on an evaluation that ranges over all possible hypotheses, or at least a set of them large enough to guarantee that it includes the true one. If we think that there is a significant chance that there is a better explanation, even one that we have not thought of, that we *cannot* think of, or that is completely unavailable to us, then we should not make the inference and normally would not. After all, an unavailable hypothesis might be the true one. It is quite unpersuasive for me to try to justify a conclusion by saying, “It is likely to be true because I couldn't think of a better explanation.” That I could not think of a better explanation is some evidence that there is no better explanation, depending on how we judge my powers of imagination and my powers of evaluating hypotheses, but these are just evidential considerations in making the judgment that there is not (anywhere) a better explanation. We should (and do) make the inferential leap when we judge that “no other hypothesis can explain  $D$  as well as  $H$  does.” Unqualified.

In one sense the *best* explanation is the true one. But, having no independent access to which explanatory hypothesis is true, the reasoner can only assert judgments of *best* based on considerations such as plausibility and explanatory power. The reasoner is, in effect, presuming that the best explanation based on these considerations is the one most likely to be true. If we challenge an abductive justification by asking what the grounds are for judging that a particular hypothesis is best, then we properly get an answer in terms of what is wrong with alternative explanations and what is the evidence that all plausible explanations have been considered (all those with a significant chance of being true). The inference schema as stated in this chapter is not trivial, even though it ranges over all possible relevant hypotheses, because *best*

is not directly a judgment of truth but instead a summary judgment of accessible explanatory virtues.

### *Emergent certainty*

Abductions often display *emergent certainty*; that is, the conclusion of an abduction can have, and be deserving of, more certainty than any of its premises. This is unlike a deduction, which is no stronger than the weakest of its links (although separate deductions can converge for parallel support). For example, I may be more sure of the bear's hostile intent than of any of the details of its hostile gestures; I may be more sure of the meaning of the sentence than of my initial identifications of any of the words; I may be more sure of the overall theory than of the reliability of any single experiment on which it is based. Patterns emerge from individual points where no single point is essential to recognizing the pattern. A signal extracted and reconstructed from a noisy channel may lead to a message, the wording of which, or even more, the intent of which, is more certain than any of its parts.

This can be contrasted with traditional empiricist epistemology, which does not allow for anything to be more certain than the observations (except maybe tautologies) since everything is supposedly built up from the observations by deduction and inductive generalization. But a pure generalization is always somewhat risky, and its conclusion is less certain than its premises. "All goats are smelly" is less certain than any given "This goat is smelly." With only deductive logic and generalization available, empirical knowledge appears as a pyramid whose base is particular experiments or sense perceptions, and where the farther up you go, the more general you get, and the less certain. Thus, without some form of certainty-increasing inference, such as abduction, traditional empiricist epistemology is unavoidably committed to a high degree of skepticism about all general theories of science.

### *Knowledge without certainty*

The conclusion of an abduction is "logically justified" by the force of the abductive argument. If the abductive argument is strong, and if one is persuaded by the argument to accept the conclusion, and if, beyond that, the conclusion turns out to be correct, then one has attained *justified, true, belief*, the classical philosophical conditions of knowledge, that date back to Plato.<sup>8</sup> Thus abductions are knowledge producing inferences despite their fallibility. Although we can never be entirely sure of an abductive conclusion, if the conclusion is indeed true, we may be said to "know" that conclusion. Of course, without independent knowledge that the conclusion is true, we do not "know that we know," but that is the usual state of our knowledge.

Summary: Abductions are fallible, and doubt cannot be completely eliminated. Nevertheless, by the aid of abductive inferences, knowledge is possible even in the face of uncertainty.

## Explanations give causes

There have been two main traditional attempts to analyze explanations as deductive proofs, neither attempt particularly successful. Aristotle maintained that an explanation is a syllogism of a certain form (Aristotle c. 350 B.C.) that also satisfies various informal conditions, one of which is that the “middle term” of the syllogism is the cause of the thing being explained. ( $B$  is the middle term of “All  $A$  are  $B$  ; All  $B$  are  $C$  ; Therefore, All  $A$  are  $C$  .”) More recently (considerably) Hempel (1965) modernized the logic and proposed the “covering law” or “deductive nomological” model of explanation.<sup>9</sup> The main difficulty with these accounts (besides Hempel’s confounding the question of what makes an ideally good explanation with the question of what it is to explain at all) is that being a deductive proof is neither necessary nor sufficient for being an explanation. Consider the following:

QUESTION: Why does he have burns on his hand?

EXPLANATION: He sneezed while cooking pasta and upset the pot.

The point of this example is that an explanation is given but no deductive proof, and although it could be turned into a deductive proof by including additional propositions, this would amount to gratuitously completing what is on the face of it an incomplete explanation. Under the circumstances (incompletely specified) sneezing and upsetting the pot were presumably *causally sufficient* for the effect, but this is quite different from being *logically sufficient*.

The case that explanations are not necessarily deductive proofs becomes even stronger if we consider psychological explanations and explanations that are fundamentally statistical (e.g., where quantum phenomena are involved). In these cases it is clear that causal determinism cannot be assumed, so the antecedent conditions cannot be assumed to be even causally sufficient for the effects. Conversely, many deductive proofs fail to be explanations of anything. For example classical mechanics is deterministic and time reversible, so an earlier state of a system can be deduced from a later state, but the earlier state cannot be said to be explained thereby. Also,  $q$  can be deduced from “ $p$  and  $q$ ” but is not thereby explained.

Thus, we conclude that explanations are not deductive proofs in any particularly interesting sense. Although they can often be presented in the form of deductive proofs, doing so does not succeed in capturing anything essential or especially useful and tends to confuse causation with logical implication.

An alternative view is that an explanation is an assignment of causal responsibility; it tells a causal story. Finding possible explanations is finding possible causes of the thing to be explained. It follows that abduction, as a

process of reasoning to an explanation, is a process of reasoning from effect to cause.

*Cause* for abduction may be understood somewhat more broadly than its usual senses of mechanical or efficient or event-event causation.<sup>10</sup> To get some idea of a more expanded view of causation, consider the four kinds of causes according to Aristotle: efficient cause, material cause, final cause, and formal cause (Aristotle, *Physics*, bk. 2, chap. 3). Let us take the example of my coffee mug. The *efficient cause* is the process by which the mug was manufactured and helps explain such things as why there are ripples on the surface of the bottom. The *material cause* is the ceramic and glaze, which compose the mug and cause it to have certain gross properties such as hardness. The *final cause* is the end or purpose, in this case to serve as a container for liquids and as a means of conveyance for drinking. A final-cause explanation is needed to explain the presence and shape of the handle. *Formal cause* is somewhat more mysterious - Aristotle is hard to interpret here - but it is perhaps something like the mathematical properties of the shape, which impose constraints resulting in certain specific other properties. That the cross-section of the mug, viewed from above, is approximately a circle, explains why the length and width of the cross-section are approximately equal. The causal story told by an abductive explanation might rely on any of these four types of causation.<sup>11</sup>

When we conclude that a finding  $f$  is explained by hypothesis  $H$ , we say more than just that  $H$  is a cause of  $f$  in the case at hand. We conclude that among all the vast causal ancestry of  $f$  we will assign responsibility to  $H$ . Typically, our reasons for focusing on  $H$  are pragmatic and connected rather directly with goals of production or prevention. We blame the heart attack on the blood clot in the coronary artery or on the high-fat diet, depending on our interests. Perhaps we should explain the patient's death by pointing out that the patient was born, so what else can you expect but eventual death? We can blame the disease on the invading organism, on the weakened immune system that permitted the invasion, or on the wound that provided the route of entry into the body. We can blame the fire on the presence of the combustibles, on the presence of the spark, or even on the presence of the oxygen, depending on which we think is the most remarkable. I suggest that it comes down to this: The things that will satisfy us as accounting for  $f$  will depend on why we are trying to account for  $f$ ; but the only things that count as candidates are parts of what we take to be the causal ancestry of  $f$ .

## Induction

Peirce's view was that induction, deduction, and abduction are three distinct types of inference, although as his views developed, the boundaries shifted somewhat, and he occasionally introduced hybrid forms such as "abductive induction" (Peirce, 1903). In this section I hope to clear up the confusion about the relationship of abduction to induction. First I argue that inductive

generalizations can be insightfully analyzed as special cases of abductions. I also argue that predictions are a distinctive form of inference, that they are not abductions, and that they are sometimes deductive, but typically not. The result is a new classification of basic inference types.

Harman (1965) argued that “inference to the best explanation” (i.e., abduction) is *the* basic form of nondeductive inference, subsuming “enumerative induction” and all other forms of nondeductive inferences as special cases. Harman argued quite convincingly that abduction subsumes sample-to-population inferences (i.e., inductive generalizations [this is my way of putting the matter]). The weakness of his overall argument was that other forms of nondeductive inference are not seemingly subsumed by abduction, most notably population-to-sample inferences, a kind of prediction. The main problem is that the conclusion of a prediction does not explain anything, so the inference cannot be an inference to a best explanation.

This last point, and others, were taken up by Ennis (1968). In his reply to Ennis, instead of treating predictions as deductive, or admitting them as a distinctive form of inference not reducible to abduction, Harman took the dubious path of trying to absorb predictions, along with a quite reasonable idea of abductions, into the larger, vaguer, and less reasonable notion of “maximizing explanatory coherence” (Harman, 1968). In this I think Harman made a big mistake, and it will be my job to repair and defend Harman’s original arguments, which were basically sound, although they proved somewhat less than he thought.

### *Inductive generalization*

First, I will argue that it is possible to treat every good (i.e., reasonable, valid) inductive generalization as an instance of abduction. An inductive generalization is an inference that goes from the characteristics of some observed sample of individuals to a conclusion about the distribution of those characteristics in some larger population. As Harman pointed out, it is useful to describe inductive generalizations as abductions because it helps to make clear when the inferences are warranted. Consider the following inference:

All observed *A* 's are *B* 's

---

Therefore All *A* 's are *B* 's

This inference is warranted, Harman (1965) writes, “. . . whenever the hypothesis that all *A* 's are *B* 's is (in the light of all the evidence) a better, simpler, more plausible (and so forth) hypothesis than is the hypothesis, say, that someone is biasing the observed sample in order to make us think that all *A*'s are *B*'s. On the other hand, as soon as the total evidence makes some other competing hypothesis plausible, one may not infer from the past correlation in the observed sample to a complete correlation in the total population.”

If this is indeed an abductive inference, then “All  $A$  's are  $B$  's” should explain “All observed  $A$  's are  $B$  's.” But, “All  $A$  's are  $B$  's” does not seem to explain why “This  $A$  is a  $B$  ,” or why  $A$  and  $B$  are regularly associated (as pointed out by Ennis, 1968). Furthermore, I suggested earlier that explanations give causes, but it is hard to see how a general fact could explain its instances, because it does not seem in any way to cause them.

The story becomes much clearer if we distinguish between an *event of observing some fact* and *the fact observed*. What the general statement in the conclusion explains is the events of observing, not the facts observed. For example, suppose I choose a ball at random (arbitrarily) from a large hat containing colored balls. The ball I choose is red. Does the fact that all of the balls in the hat are red explain why this particular ball is red? No. But it does explain why, when I chose a ball at random, it turned out to be a red one (because they all are). “All  $A$  's are  $B$  's” cannot explain why “This  $A$  is a  $B$  ” because it does not say anything at all about how its being an  $A$  is connected with its being a  $B$  . The information that “they all are” does not tell me anything about why this one is, except it suggests that if I want to know why this one is, I would do well to figure out why they all are.

A generalization helps to explain the events of observing its instances, but it does not explain the instances themselves. That the cloudless, daytime sky is blue helps explain why, when I look up, I see the sky to be blue (but it doesn't explain why the sky is blue). The truth of “Theodore reads ethics books a lot” helps to explain why, so often when I have seen him, he has been reading an ethics book (but it doesn't explain why he was reading ethics books on those occasions). Seen this way, inductive generalization does have the form of an inference whose conclusion explains its premises.

Generally, we can say that *the frequencies in the larger population, together with the frequency-relevant characteristics of the method for drawing a sample, explain the frequencies in the observed sample*. In particular, “ $A$  's are mostly  $B$  's” together with “This sample of  $A$  's was drawn without regard to whether or not they were  $B$  's” explain why the  $A$  's that were drawn were mostly  $B$  's.

Why were 61% of the chosen balls yellow?

Because the balls were chosen more or less randomly from a population that was two thirds yellow (the difference from  $2/3$  in the sample being due to chance).

Alternative explanation for the same observation:

Because the balls were chosen by a selector with a bias for large balls from a population that was only one third yellow but where yellow balls tend to be larger than non yellow ones.

How do these explain? By giving a causal story.

What is explained is (always) some *aspect* of an event/being/state, not a whole event/being/state itself. In this example just the frequency of characteristics in the sample is explained, not why these particular balls are yellow or why the experiment was conducted on Tuesday. The explanation explains why the sample frequency was the way it was, rather than having some markedly different value. In general, if there is a deviation in the sample from what you would expect, given the population and the sampling method, then you have to throw some Chance into the explanation (which is more or less plausible depending on how much chance you have to suppose).<sup>12</sup>

The objects of explanation - what explanations explain - are facts about the world (more precisely, always an aspect of a fact, under a description). Observations are facts; that is, an observation having the characteristics that it does is a fact. When you explain observed samples, an interesting thing is to explain the frequencies. A proper explanation will give a causal story of how the frequencies came to be the way they were and will typically refer both to the population frequency and the method of drawing the samples.

Unbiased sampling processes tend to produce representative outcomes; biased sampling processes tend to produce unrepresentative outcomes. This “tending to produce” is causal and supports explanation and prediction. A peculiarity is that characterizing a sample as “representative” is characterizing the effect (sample frequency) by reference to part of its cause (population frequency). Straight inductive generalization is equivalent to concluding that a sample is representative, which is a conclusion about its cause. This inference depends partly on evidence or presumption that the sampling process is (close enough to) unbiased. The unbiased sampling process is part of the explanation of the sample frequency, and any independent evidence for or against unbiased sampling bears on its plausibility as part of the explanation.

If we do not think of inductive generalization as abduction, we are at a loss to explain why such an inference is made stronger or more warranted, if in collecting data we make a systematic search for counter-instances and cannot find any, than it would be if we just take the observations passively. Why is the generalization made stronger by making an effort to examine a wide variety of types of *A* 's? The inference is made stronger because the failure of the active search for counter-instances tends to rule out various hypotheses about ways in which the sample might be biased.

In fact the whole notion of a “controlled experiment” is covertly based on abduction. What is being “controlled for” is always an alternative way of explaining the outcome. For example a placebo-controlled test of the efficiency of a drug is designed to make it possible to rule out purely psychological explanations for any favorable outcome.

Even the question of sample size for inductive generalization can be seen clearly from an abductive perspective. Suppose that on each of the only two occasions when Konrad ate pizza at Mario's Pizza Shop, he had a stomachache the next morning. In general, Konrad has a stomachache occasionally but not frequently. What may we conclude about the relationship between the pizza and the stomachache? What may we reasonably predict about the outcome of

Konrad's next visit to Mario's? Nothing. The sample is not a large enough. Now suppose that Konrad continues patronizing Mario's and that after every one of 79 subsequent trips he has a stomach ache within 12 hours. What may we conclude about the relationship between Mario's pizza and Konrad's stomachache? That Mario's pizza makes Konrad have stomachaches. We may predict that Konrad will have a stomachache after his next visit, too.

A good way to understand what is occurring in this example is by way of abduction. After Konrad's first two visits we could not conclude anything because we did not have enough evidence to distinguish between the two competing general hypotheses:

1. The eating pizza - stomachache correlation was accidental (i.e., merely coincidental or spurious [say, for example, that on the first visit the stomach ache was caused by a virus contracted elsewhere and that on the second visit it was caused by an argument with his mother]).
2. There is some connection between eating pizza and the subsequent stomach ache (i.e., there is some causal explanation of why he gets a stomach ache after eating the pizza [e.g., Konrad is allergic to the snake oil in Mario's Special Sauce]).

By the time we note the outcome of Konrad's 79th visit, we are able to decide in favor of the second hypothesis. The best explanation of the correlation has become the hypothesis of a causal connection because explaining the correlation as accidental becomes rapidly less and less plausible the longer the association continues.

### *Prediction*

Another inference form that has often been called “induction” is given by the following:

All observed *A*'s are *B*'s.

---

Therefore, the next *A* will be a *B*.

Let us call this inference form an *inductive projection*. Such an inference can be analyzed as an inductive generalization followed by a prediction, as follows:

Observations ----> All *A*'s are *B*'s ----> The next *A* will be a *B*.

Predictions have traditionally been thought of as deductive inferences. However, something is wrong with this analysis. To see this, consider the alternative analysis of inductive projections, as follows:

Observations  $\rightarrow$  At least generally  $A$ 's are  $B$ 's  $\rightarrow$  The next  $A$  will be a  $B$ .

This inference is stronger in that it establishes its conclusion with more certainty, which it does by hedging the generalization and thus making it more plausible, more likely to be true. It could be made stronger yet by hedging the temporal extent of the generalization:

Observations  $\rightarrow$  At least generally  $A$ 's are  $B$ 's, at least for the recent past and the immediate future  $\rightarrow$  The next  $A$  will be a  $B$ .

The analyses of inductive projection with the hedged generalizations are better than the first analysis because they are better at making sense of the inference, which they do by being better at showing the sense in it (i.e., they are better at showing how, and when, and why the inference is justified - or "rational" or "intelligent"). Reasonable generalizations are hedged. Generally the best way to analyze " $A$ 's are  $B$ 's" is not "All  $A$ 's are  $B$ 's," as we are taught in logic class, but as "Generally  $A$ 's are  $B$ 's," using the neutral, hedged, universal quantifier of ordinary life.<sup>13</sup>

We have analyzed inductive projections as inductive generalizations followed by predictions. The inductive generalizations are really abductions, as was argued before. But, what kind of inferences are predictions? One thing seems clear: *Predictions from hedged generalizations are not deductions.*

Predictions from hedged generalizations belong to the same family as *statistical syllogisms* which have forms like these:<sup>14</sup>

$m/n$  of the  $A$ 's are  $B$ 's (where  $m/n > 1/2$ ).

---

Therefore, the next  $A$  will be a  $B$ .

and

$m/n$  of the  $A$ 's are  $B$ 's.

---

Therefore, approximately  $m/n$  of the  $A$ 's in the next sample will be  $B$ 's.

These are also related to the following forms:

Generally  $A$ 's are  $B$ 's.  
 $S$  is an  $A$ .

---

Therefore,  $S$  is a  $B$ .

and

A typical, normal  $X$  does  $Y$ .

---

Therefore, this  $X$  will do  $Y$ .

None of these inferences appear to be deductions. One can, of course, turn

them into deductions by including a missing premise like “this  $X$  is normal,” but unless there is independent evidence for the assumption, this adds nothing to the analysis.

Furthermore, consider an inference of this form:

$P$  has high probability.

---

Therefore,  $P$ .

Such an inference (whatever the probability is taken to mean) will have to allow for the possibility of the conclusion being false while the premise is true. This being so, such an inference cannot possibly be deductive.

Thus it seems that some of our most warranted inductive projections, those mediated by various kinds of hedged generalizations, do not rely on deduction for the final predictive step. These predictions, including statistical syllogisms, are not deductions, and they do not seem to be abductions. That is, *sometimes* the thing predicted is also explained, but note that the conclusions in abductions do the explaining, whereas for predictions, if anything, the conclusions are what is explained. Thus the predictive forms related to statistical syllogism are in general nondeductive and nonabductive as well.

If predictions are not abductions, what then is the relationship between prediction and explanation? The idea that they are closely related has a fairly elaborate history in the philosophy of science. Some authors have proposed that explanations and predictions have the same logical form. Typically this is given as the form of a proof whereby the thing to be explained or predicted is the conclusion, and causation enters in to the premises somehow, either as a causal law or in some other way.

The idea seems to be that to explain something is to be in a position to have predicted it, and to predict something puts one in a position to explain it, if it actually occurs. This bridges the apparent basic asymmetry that arises because what you explain is more or less a given (i.e., has happened or does happen), whereas what you predict is an expectation and (usually) has not already happened.

Despite its apparent plausibility, this thesis is fundamentally flawed. There is no necessary connection between explanation and prediction of this sort. Consider the following two counterexamples.

Example 1. *George-did-it explanations*. Why is there mud on the carpet?

Explanation: George did it (presumably, but not explicitly, by tracking it into the house stuck to his shoes, or something similar). Knowing that George came into the room puts me in a position to predict mud on the carpet only if I assume many questionable auxiliary assumptions about George’s path and the adhesive properties of mud, and so forth. If I had an ideally complete explanation of the mud on the carpet, some sort of complete causal story, then, perhaps, I would

be in a position to predict the last line of the story, given all the rest; but this is an unrealistic representation of what normal explanation is like. George-did-it explanations can be perfectly explanatory without any pretensions of completeness. This shows that *one can explain without being in a position to predict*. Many other examples can be given where explanation is ex post facto, but where knowledge of a system is not complete enough to license prediction. Why did the dice come up double 6s? Because of throwing the dice, and chance. We are often in a position to explain a fact without being in a position to have predicted it - specifically, when our explanation is not complete, which is typical, or when the explanation does not posit a deterministic mechanism, which is also typical.

*Example 2. Predictions based on trust.* Suppose that a mechanic, whom I have good reason to trust, says that my car will soon lose its generator belt. On this basis I predict that the car will lose its generator belt (on a long drive home, say). Here I have made a prediction, and on perfectly good grounds, too, but I am not in a position to give an explanation (I have no idea what has weakened the belt). This example, weather predictions, and similar examples of predictions based on authority show that *one can be in a position to predict without being in a position to explain*.

I believe that explanations are causal, and that predictions are commonly founded on projecting consequences based on our causal understanding of things. Thus, commonly, an explanation of some event  $E$  refers to its causes, and a prediction of  $E$  is based on its causes, and both the explanation and the prediction suppose the causal connections. However, I believe that the symmetry between explanation and prediction goes no further.

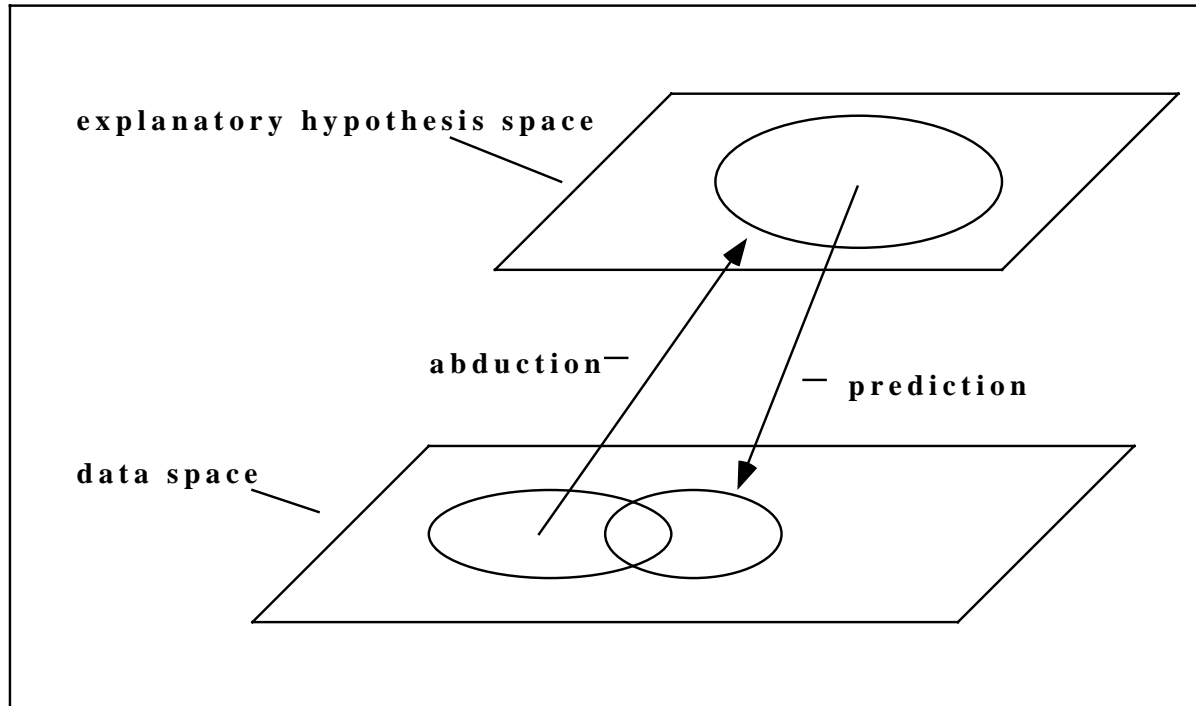
When a prediction fails, it casts doubt on the general premises on which it is based. This is part of the logical underpinnings of scientific reasoning. The view presented here is similar to what has been called the “hypothetico-deductive” model of scientific reasoning, except in insisting that hypotheses must be explanatory, and in denying that predictions are always deductive.

Predictions, then, are neither abductions nor (typically) deductions. This is contrary, both to the view that predictions are deductions and to Harman's view that all nondeductive inferences are abductions. Rather, predictions and abductions are two distinct kinds of plausible inference. Abductions go from data to explanatory hypothesis; predictions go from hypothesis to expected data. (See Figure 1.1.)

Jerry Hobbs has suggested (verbally) that, “The mind is a big abduction machine.” In contrast Eugene Charniak has suggested (verbally) that there are two fundamental operations of mind: abduction and planning. The view presented in this chapter, while close to that of Hobbs in its enthusiasm for abduction, is actually closer to Charniak's. It elaborates that view, however, by adding that planning depends on prediction (to anticipate consequences of actions), and it is prediction that is inferentially fundamental. Planning is

choosing actions based on expected outcomes. So planning is “reasoning” all right, but it is not “inference,” since planning decides action rather than belief.

While asserting that abduction and prediction are inferentially distinct, we note that they are often entangled as processes. Sometimes an abduction will use prediction as a subtask (e.g., for testing a hypothesis), and sometimes a prediction will use abduction as a subtask (e.g., for assessing the situation).



**Figure 1.1. Abduction and prediction.**

### *Probabilities and abductions*

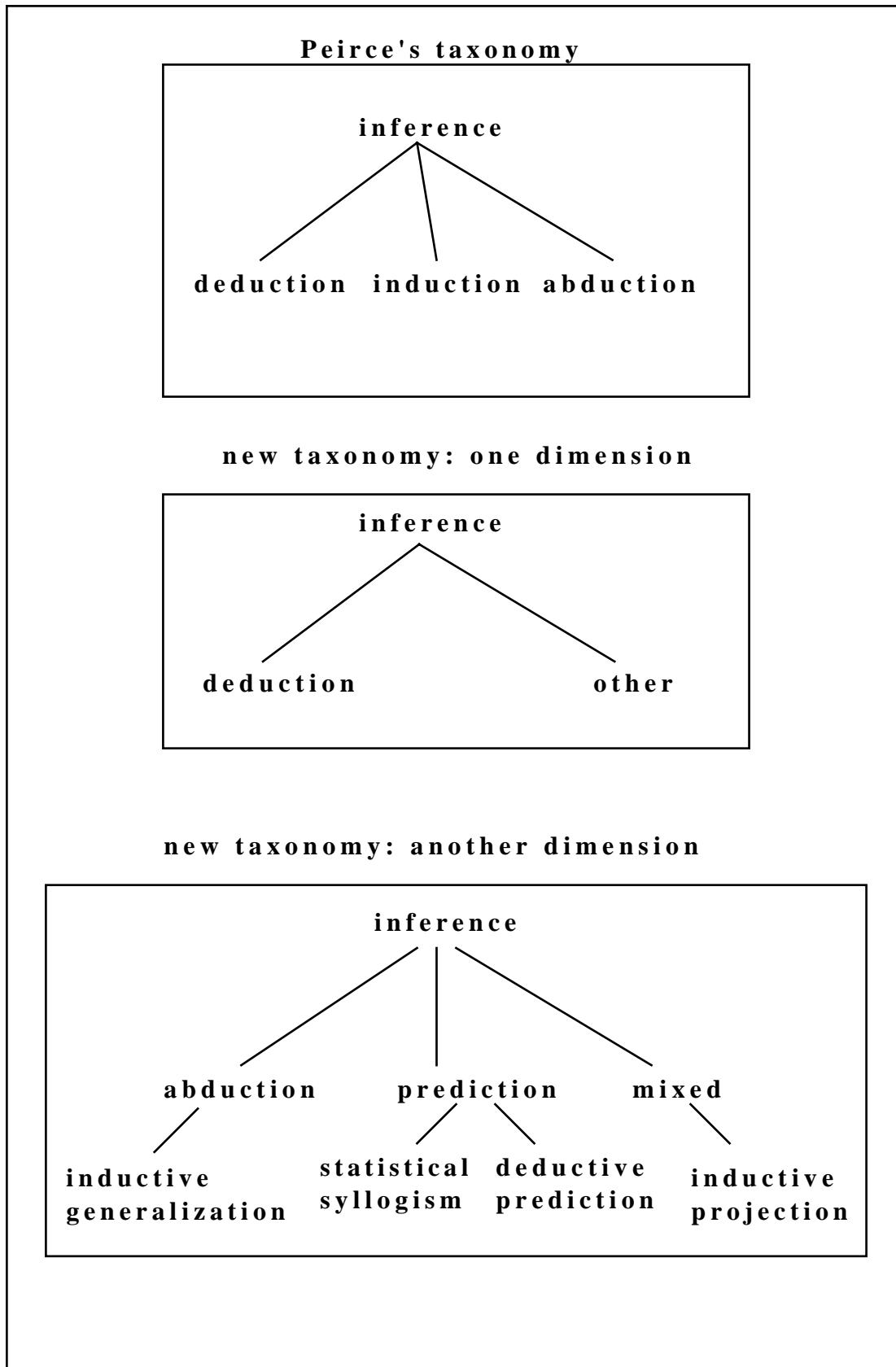
It has been suggested that we should use mathematical probabilities to help us choose among explanatory hypotheses. (Bayes's Theorem itself can be viewed as a way of describing how simple alternative causal hypotheses can be weighed.) If suitable knowledge of probabilities is available, the mathematical theory of probabilities can, in principle, guide our abductive evaluation of explanatory hypotheses to determine which is best. However, in practice it seems that rough qualitative confidence levels on the hypotheses are enough to support abductions, which then produce rough qualitative confidence levels for their conclusions. It is certainly possible to model these confidences as numbers from a continuum, and on rare occasions one can actually get knowledge of numerical confidences (e.g., for playing blackjack). However, for the most part numerical confidence estimates are unavailable and unnecessary for reasoning. People are good abductive reasoners without close

estimates of confidence. In fact it can be argued that, if confidences need to be estimated closely, then it must be that the best hypothesis is not much better than the next best, in which case no conclusion can be confidently drawn because the confidence of an abductive conclusion depends on how decisively the best explanation surpasses the alternatives. Thus it seems that confident abductions are possible only if confidences for hypotheses do not need to be estimated closely.

Moreover, it appears that accurate knowledge of probabilities is not commonly available because the probability associated with a possible event is not very well defined. There is almost always a certain arbitrariness about which reference class is chosen as a base for the probabilities; the larger the reference class, the more reliable the statistics, but the less relevant they are; whereas the more specific the reference class, the more relevant, but the less reliable. (See Salmon, 1967, p. 92.) Is the likelihood that the next patient has the flu best estimated based on the frequency in all the people in the world over the entire history of medicine? It seems better at least to control for the season and to narrow the class to include people at just this particular time of the year. (Notice that causal understanding is starting to creep into the considerations.) Furthermore, each flu season is somewhat different, so we would do better to narrow to considering people just *this* year. Then, of course, the average *patient* is not the same as the average *person*, and so forth, so the class should probably be narrowed further to something such as this: people of this particular age, race, gender, and social status who have come *lately* to doctors of this sort. Now the only way the doctor can have statistics this specific is to rely on his or her own most recent experience, which allows for only rough estimates of likelihood because the sample is so small. There is a Heisenberg-like uncertainty about the whole thing; the closer you try to measure the likelihoods, the more approximate the numbers become. In the complex natural world the long-run statistics are often overwhelmed by the short-term trends, which render the notion of a precise prior probability of an event inapplicable to most common affairs.

### *Taxonomy of basic inference types*

Considering its apparent ubiquity, it is remarkable how overlooked and underanalyzed abduction is by almost 2,400 years of logic and philosophy. According to the analysis given here, the distinction between abduction and deduction is a distinction between different dimensions, so to speak, of inference. Along one dimension inference can be distinguished into deductive and nondeductive inference; along another dimension inferences can be distinguished as abductive and predictive (and mixed) sorts of inferences. Abduction absorbs inductive generalization as a subclass and leaves the predictive aspect of induction as a separate kind of inference. Statistical syllogism is a kind of prediction. This categorization of inferences is summarized in Figure 1.2.



**Figure 1.2. Taxonomy of basic inference types.**

## From wonder to understanding

Learning is the acquisition of knowledge. One main form of learning starts with wonder and ends in understanding. To understand something is to grasp an explanation of it. (To explain something to somebody is to package up an understanding and communicate it.) Thus knowledge is built up of explanatory hypotheses, which are put in place in memory as a result of processes set in motion by wondering “Why?” That is, one main form of knowledge consists of answers to explanation-seeking why questions.

An explanation-seeking why question rests on a Given, a presupposition upon which the question is based. For example, “Why is the child sneezing?” presupposes that the child is indeed sneezing. This Given is not absolutely firm, even though it is accepted at the outset, for in the end we may be happy to throw it away, as in, “Oh, those weren't sneezes at all. She was trying to keep a feather in the air by blowing.” Usually, perhaps always, along with the Given some contrasting possibility is held in mind, some imagined way that the Given could have been different (Bromberger, 1966). Thus, behind the *G* in a “Why *G*?” usually there appears an “. . . as opposed to *H*” discernible in the background.

Abduction is a process of going from some Given to a best explanation for that (or related) given. Describing a computational process as abduction says *what* it accomplishes - namely, generation, criticism, and acceptance of an explanation - but superficially it says nothing about *how* it is accomplished (for example, how hypotheses are generated or how generation interacts with acceptance).

An explanation is an assignment of causal responsibility; it tells a causal story (at least this is the sense of “explanation” relevant to abduction). Thus, finding possible explanations is finding possible causes of the thing to be explained, and so abduction is essentially a process of reasoning from effect to cause.

A large part of knowledge consists of causal understanding. Abductions produce knowledge, both in science and in ordinary life.

# Index

## A

abduction  
 and deduction, 9–10  
 and prediction, 23  
 and probabilities, 23–24  
 as a pattern of evidential relationships, 9  
 as a pattern of justification, 6  
 as a subtask of prediction, 23  
 as part of logic, 9  
 as reasoning from effect to cause, 26  
 characterization of, 1, 11, 26  
 conclusion, 13  
 deliberate, 2  
 descriptive theory of, 12  
 evaluative theory of, 12  
 explanatory theory of, 12  
 fallibility of, 13  
 goal of, 10  
 in historical scholarship, 4  
 in language understanding, 2, 5  
 in ordinary life, 2, 26  
 in perception, 2  
 in science, 4, 26  
 includes generation and possible acceptance, 5  
 normative theory of, 12  
 perceptual, 2  
 suspended, 11  
 task definition, 5  
 abductive justification, 6–9  
 ampliative inference, 10  
 antecedent conditions, 14  
 Aristotle, 14, 15  
 Aristotle. note 1.11

## B

Bayes's Theorem, 23  
 belief, 9  
 best explanation  
 compared to what?, 12–13  
 Bhaskar, R.. note 1.9  
 biased sampling processes, 18  
 Bonaparte, N., 4  
 Bordley, J., 6, 8  
 Bromberger, S., 26

## C

causal responsibility, 14, 15, 26  
 causal sufficiency, 14  
 causal understanding, 26  
 causality  
 and explanation. note 1.10  
 and implication, 14  
 causation, 14–15  
 cause  
 efficient, 15  
 final, 15  
 formal, 15

material, 15  
 mechanical, 15  
 cause, 14  
 certainty, 13  
 emergent, 13  
 increased by hedging, 20  
 certainty-increasing inference, 13  
 chance, as explanatory, 18  
 Charniak, E., 1, 22  
 coherence, 16  
 confidence  
 coarse scale, 23  
 numerical, 23  
 contrasting possibility, 26  
 controlled experiment, 18  
 correlation, 19  
 correlation, 16  
 criticism, 26  
 crucial experiment, 3

## D

Darden, L.. note 1.3  
 Darwin, C., 4  
 data gathering, 11, 18  
 deception, 11  
 deduction, 13  
 and abduction, 10  
 and explanation, 14  
 and prediction, 20  
 Descartes, R., 10  
 detective, 4  
 determinism, 14  
 diagnosis, 6–9  
 errors in, 7–8  
 diagnosis, 6  
 diagnostic dialog, 8  
 diagnostic differential, 6  
*disjunctive syllogism*, 9  
 Donahue, T.. note 1.12  
 doubt, 10  
 cannot be eliminated, 11  
 persistence of, 13  
 Doyle, A. C., 4

## E

*emergent certainty*, 13  
 empirical knowledge, 13  
 empiricism, 13  
 Ennis, R., 16, 17  
 enumerative induction, 16  
 evidence  
 negative to positive. note 1.5  
 evidential relationships, science of, 9  
 evidential support, 9  
 expectations  
 failure of, 18  
 experiment, 3, 13, 18  
 explanandum, 18  
 explanation, 14–15, 26  
 and deductive proof, 14  
 and pragmatics, 15  
 and prediction, 21–22

and understanding, 26  
 best, 12  
 by general statements, 17  
 cognitive aspects of, 15  
 covering-law model of, 17  
 deductive-nomological model of, 14, 17  
 ideally complete, 21  
 incomplete, 14  
 models of, 14  
 object of, 18, 26  
 of correlation, 19  
 of sample frequencies, 16–19  
 plausible, 12  
 psychological, 14  
 statistical, 14  
 true, 7, 12  
 explanation, 14. note 1.9  
 explanations  
 George-did-it type, 21  
 explanatory inference, 1

## F

fair questions, 8  
 Fann, K. T., 4

## G

generalization, hedged, 20  
 Gettier problem. note 1.8  
 givens, 26

## H

Harman, G., 1, 2, 3, 4, 16, 22  
 Harman, G.. note 1.8  
 Harvey, A., 6  
 Hastie, R., 1  
 hedged generalizations, 20  
 hedging, to strengthen inference, 20  
 Hempel, C., 14  
 hesitation, 10  
 Hobbs, J., 22  
 Holmes, S., 4  
 hypothesis  
 acceptance, 5  
 composite, 5  
 criticism, 5  
 generation, 5  
 hypothetico-deductive model, 22

## I

implication, causality and, 14  
 induction, 19, 15–24  
 inductive generalization, 13, 16–19  
*inductive projection*, 19, 21  
 inference, 23  
 ampliative, 10  
 defined, 9  
 fallible, 9  
 function of, 9

strengthened by hedging, 20  
 taxonomy of types, 24  
 truth preserving, 10  
 truth producing, 10  
 inference to the best explanation,  
 see abduction  
 information-seeking processes, 11  
 intelligence, 12

**J**

Josephson, J., 1  
 jury, 1, 3  
 justified true belief, 13

**K**

knowing that we know, 13  
 knowledge  
 and explanatory hypotheses,  
 26  
 in the philosophical sense, 13  
 justified true belief, 13  
 possibility of, 13  
 knowledge producing inferences,  
 13

**L**

law, 3  
 learning, 26  
 likelihood, 11  
 logic  
 as the science of evidential  
 relationships, 9  
 logic of discovery. note 1.6  
 logic of justification. note 1.6  
 logical implication, confused with  
 causation, 14  
 low-likelihood alternative  
 explanations, 11  
 Lycan, W., 1  
 Lycan, W.. note 1.7. note 1.1

**M**

mathematical logic, weakness of.  
 note 1.13

maximizing explanatory  
 coherence, 16  
 McDermott, D., 1  
 medical diagnosis, 1  
 Milne, A. A., 3  
 mind, fundamental operations of,  
 22  
 modus ponens, 1

**N**

Newton, I., 4  
 noise hypothesis, 5  
 noisy channel, 13  
 normality, 6, 21

**O**

observation, 13, 16, 17, 18  
 observation language, 10  
 other minds, 2

**P**

Peirce, C. S., 1, 4, 15  
 Peng, Y., 6  
 Pennington, N., 1  
 placebo, 18  
 planning, 22  
 Plato, 13  
 plausibility, 10  
 Pooh, W. T., 3  
 population-to-sample inference,  
 16  
 prediction, 16, 19–24  
 and explanation, 21–22  
 as a subtask of abduction, 23  
 based on trust, 22  
 failure of, 22  
 is not deduction, 20  
 probability, 21  
 and abduction, 23–24  
 prior, 24  
 problem of induction, 19–21

**Q**

quantifier of ordinary life, 20

**R**

rationality  
 and hedging, 20  
 reference-class problem, 24  
 Reggia, J., 6  
 revising the data, 5

**S**

Salmon, W., 24  
 Salmon, W.. note 1.9  
 sample size, 18  
 sample, biased, 16, 17, 18  
 sample-to-population inference,  
 16  
 sampling processes, 18  
 Sebeok, T., 4  
 skepticism, 13  
 spurious, 19  
*statistical syllogism*, 20

**T**

Tanner, M., 1  
 teleology, 15  
 term introduction, 10  
 testimony, 3  
 Thagard, P.. note 1.2  
 theory language, 10  
 truth preservation, 10  
 truth production, 10  
 Truzzi, M., 4

**U**

Umiker-Sebeok, J., 4  
 understanding, 26  
 universal quantifier, 20

**W**

Wallace, W.. note 1.10  
 why questions, 26  
 witness, 3  
 wonder, 26

## Notes

---

<sup>1</sup> This formulation is largely due to William Lycan.

<sup>2</sup> Thagard (1988) recognizes abduction in his analysis of scientific theory formation.

<sup>3</sup> For example, Darden (1991) describes the modularity of genetic theory.

<sup>4</sup> The remainder of this chapter is more philosophical. It is not necessary to accept everything in order to find value in the rest of the book. The next chapter includes an orientation to our approach to AI and a discussion of representational issues. The main computational treatment of abduction begins in chapter 3. The philosophy increases again at the end of the book.

<sup>5</sup> Thus abductions have a way of turning negative evidence against some hypotheses into positive evidence for alternative explanations.

<sup>6</sup> This condition shows most clearly why the process of discovery is a logical matter, and why logic cannot simply be confined to matters of justification. The “logic of justification” cannot be neatly separated from the “logic of discovery” because justification depends, in part, on evaluating the quality of the discovery process.

<sup>7</sup> Suggested by William Lycan.

<sup>8</sup> I am ignoring here the so called “Gettier problem” with these conditions, but see Harman (1965) for an argument that successful abductions resolve the Gettier problem anyway.

<sup>9</sup> For a brief summary of deductive and other models of explanation see Bhaskar (1981), and for a history of recent philosophical accounts of explanation, see Salmon (1990).

<sup>10</sup> For a well-developed historical account of the connections between ideas of causality and explanation see Wallace (1972, 1974). Ideas of causality and explanation have been intimately linked for a very long time.

<sup>11</sup> What the types of causation and causal explanation are remains unsettled, despite Aristotle’s best efforts and those of many other thinkers. The point here is that a narrow view of causation makes progress harder by obscuring the degree to which all forms of causal thinking are fundamentally similar.

<sup>12</sup> “It is embarrassing to invoke such a wildly unlikely event as a chance encounter between the entry probe and a rare and geographically confined methane plume, but so far we have eliminated all other plausible explanations” (Planetary scientist Thomas M. Donahue of the University of Michigan on the analysis of chemical data from a Pioneer probe parachuted onto the planet Venus, reported in *Science News* for Sept. 12, 1992).

<sup>13</sup> Note that this analysis suggests an explanation for why traditional mathematical logic has been so remarkably unsuccessful in accounting for reasoning outside mathematics and the highly mathematical sciences. The universal quantifier of logic is not the universal quantifier of ordinary life, or even of ordinary scientific thought.

<sup>14</sup> I have not put likelihood qualifiers in the conclusions of any these forms because doing so would at best postpone the deductive gap.