Brit. J. Phil. Sci. **54** (2003), 353–358, axg019

REVIEW

LORENZO MAGNANI

Abduction, Reason, and Science: Processes of Discovery and Explanation

Kluwer Academic Publishers/Plenum Publishers, 2000; cloth $85
ISBN: 0-306-46514-0

Jon Williamson
King's College
University of London

What distinguishes abductive inference from other modes of inference? What distinguishes the different forms of abductive inference? These are perhaps the key questions that face the student of abduction, and the core of Lorenzo Magnani's book is devoted to them. The book contains, among other things, a defence of several putative distinctions, and it touches on an impressively wide range of perspectives, citing much of the relevant work in artificial intelligence and cognitive science.

Alas, abductive inference is curiously resistant to sharp distinctions, as I shall attempt to explain.

C. S. Peirce used the term 'abduction' to refer to the process of reasoning to explanation:

> Accepting the conclusion that an explanation is needed when facts contrary to what we should expect emerge, it follows that the explanation must be such a proposition as would lead to the prediction of the observed facts, either as necessary consequences or at least as very probable under the circumstances. A hypothesis then, has to be adopted which is likely in itself, and renders the facts likely. This step of adopting a hypothesis as suggested by the facts is what I call *abduction*.[1]

The question naturally arises as to how abduction is to be distinguished from other types of reasoning. Peirce maintains that,

> Abduction is the process of forming an explanatory hypothesis [. . .] Deduction proves that something *must* be; Induction shows that something *actually is* operative; Abduction merely suggests that something *may be*.[2]

Moreover, abduction's 'only justification is that from its suggestion deduction can draw a prediction which can be tested by induction, and

[1] [Peirce 1932–1963] §7.202.
[2] [Peirce 1932–1963] §5.171.

that, if we are ever to learn anything and understand phenomena at all, it must be by abduction that this is to be brought about.'[3] Magnani elaborates this view by putting forward what he calls the ST-MODEL: abduction is the process of coming up with a hypothesis; deduction is used to draw the consequences of the hypothesis, and induction evaluates the hypothesis on the basis of whether or not its consequences occur.

Unfortunately this distinction between the roles of abduction, deduction and induction is not altogether a happy one. Two of the difficulties are terminological. First, Peirce was not altogether clear whether abduction refers to the first part of the process—conjecturing a hypothesis—or to the whole process—conjecturing, drawing consequences and evaluation. As Peirce acknowledges, 'Concerning the relations of these three modes of inference [. . .] my opinions, I confess, have wavered.'[4] Second, induction is often understood to be the process of generalisation, as well as that of evaluation or confirmation. Thus construed, induction is a type of abduction, where the required explanation is a generalisation of some of the initial data.[5]

The third difficulty is more fundamental. There are normally too many possible explanatory hypotheses to conjecture and evaluate each one in turn.[6] Rather, one must conjecture a few *plausible* hypotheses, and evaluate these—if any are good explanations one might choose the best, otherwise one can conjecture again. Thus the process of evaluation is intertwined with that of conjecturing: one must only conjecture hypotheses which are likely to evaluate well. This blurs the distinction between abduction and induction.

Magnani claims that the distinction remains tenable in the context of artificial reasoning systems, where one can artificially separate the processes of conjecturing and evaluation. I doubt this: a space of possible hypotheses is rarely small enough to allow exhaustive search; normally one must use the data to guide search, i.e. one must somehow evaluate the unarticulated hypotheses in the search space in order to entertain one or more of them as conjectures. Conjecturing and evaluation are concurrent. Moreover, the empirical successes of data-constrained search in artificial intelligence systems can be used to argue against the usefulness of a strategy of conjecturing independently of the data, even if such a strategy were to be feasible in practice.[7] Note that some AI proposals use data to constrain search for

[3] [Peirce 1932–1963] §5.171.

[4] [Peirce 1932–1963] §5.146. Thagard ([1988]) §4.2.1 argues that Peirce first viewed abduction as the whole process of reasoning to explanation, only later viewing it as just conjecturing, and that the editors of Peirce's collected papers have obscured this shift in ideas.

[5] As Peirce points out, some 'inferences, which are really inductions, sometimes present, nevertheless, some indubitable resemblances to hypotheses' (Peirce [1932–1963] §2.636).

[6] 'If he examines all the foolish theories he might imagine, he never will (short of a miracle) light upon the true one' Peirce [1932–1963], §2.776.

[7] Some such successes are cited in Gillies ([1996]).

plausible conjectures yet also claim to separate the processes of conjecturing and evaluation. Typically, they suggest the use of logic programming to determine a set of plausible hypotheses that yield the data as a consequence, and then the use of statistics to evaluate each of the hypotheses in the light of the data.[8] But the claim is false—evaluation has not been separated from conjecturing. In these systems logic is used as a preliminary evaluation mechanism to allow conjectures to be formulated, and then probability is used to refine the evaluation. There is no blind conjecturing in these systems: all the mooted hypotheses are confirmed by the data in as much as they imply the data.

On the other hand, Magnani does reject a sharp distinction between conjecturing and evaluation in the context of describing human reasoning. Here he cites empirical evidence to the effect that expert physicians use initial data to constrain the conjecturing process, while intermediate and novice physicians conjecture first, and then draw consequences and evaluate in the light of initial data. In this instance, reasoning to explanation is significantly more successful when the conjecturing and evaluation phases are combined.

Magnani distinguishes several different types of abduction. First and foremost in his classification is the distinction between theoretical and manipulative abductions. Although it is never made entirely explicit, it appears that in the case of theoretical abduction, explanations are representational entities, such as sentences and pictures, while in the case of manipulative abduction, explanations are objects with which we interact or the interactive processes themselves. While theoretical abduction requires the formulation of hypotheses which say something about the world, manipulative abduction, on the other hand, involves manipulating objects in experiments, or simulated experiments. For Magnani, both types of abduction are reasoning processes in their own right: 'theoretical abduction is the process of *inferring* certain facts and/or laws and hypotheses that render some sentences plausible, that *explain* or *discover* some (eventually new) phenomenon or observation' (p. 17) while '*Manipulative* abduction [. . .] happens when we are thinking *through* doing and not only, in a pragmatic sense, about doing' (p. 53).

But are experimenting and theorising really distinct forms of reasoning from explanation? As Magnani acknowledges, experiments often allow one to get a feel for a problem, and are thus a precursor to formulating a theory. But of course experiments are also used to test a theory and to suggest improvements to it. Thus, just as conjecturing and evaluation are concurrent aspects of the same process, so too experiment and theorising are inseparable.

8 See Flach & Kakas ([2000]) for example.

They are not different types of abduction: they are mutually supportive activities that together lead to successful abductive inference.[9]

Magnani's classification continues by distinguishing two types of theoretical abduction, sentential and model-based. In sentential abduction, explanations are linguistic entities, such as sentences of a logical language, frames, and neural and probabilistic networks. Model-based abduction covers non-sentential representations, such as diagrams, images and memories. Magnani considers model-based abduction to be less formal than sentential abduction, and to account for perceptual, causal and analogical reasoning. Again, one can cast doubt on the sharpness of such a distinction: what separates visual or pictorial representations from other types of representation? Indeed, many argue that perceptual inference can be accounted for within a neural network framework, causal reasoning can be accounted for within a probabilistic network framework, and analogical reasoning can be accounted for within a frame-based representation.

A further distinction is made between selective and creative abduction. In selective abduction a hypothesis is chosen from a well-defined hypothesis space, whereas in creative abduction the conjectured hypothesis is new in some more radical way. By way of example, Magnani contrasts medical diagnosis, where a hypothesis is selected from all those known, and medical research, where new diseases and causal relations are articulated and confirmed.

Magnani maintains that the sentential/model-based distinction aligns with the selective/creative distinction: sentential abduction is selective while model-based abduction is creative. The line of argument appears to be something like this: artificial intelligence systems employ a search approach and these systems fit into the sentential framework; human explanatory reasoning is non-linguistic or involves radical changes in language, and so is model-based, but also often creative.

In my view there is little substance to the selective/creative distinction. I will be bold: all reasoning to explanation is selective. All abductive inference can be thought of as selecting a conjecture in the search space of all hypotheses. Normally we search locally, from current hypotheses to those with minor differences, but occasionally we jump to new parts of the search space, by, for example, dropping closely-held assumptions or altering the language in which hypotheses are expressed, and it is these transitions that are considered creative. We do not represent the entire search space—we could not, because it is too big—but we often articulate hypotheses in the

[9] Any sharp distinction between experiment and theory certainly sits at odds with the accounts in Galison ([1997]), for instance, which shows how experiments are thoroughly integrated into the practice of scientific theorising.

local vicinity in advance and use data and heuristics to select a direction in which to search.

Take the medical example. In a typical diagnosis problem, a specialist or artificial expert system might have a representation of the causal relations in her area of specialism, and her explanatory task is to ascertain a patient's problematic symptoms and decide what is causing them. We can represent the causal knowledge as a directed graph linking variables of interest, and then the task is to select the most likely values of those variables for a given patient. A common recommendation from artificial intelligence is to devise a Bayesian network by augmenting the causal graph with the probability distribution of each node conditional on its direct causes, and use this Bayesian network to calculate the most likely values of variables given the observations made.

From the point of view of the medical statistician whose job it is to ascertain the conditional probability distributions as parameters of the Bayesian network, the task is to use past patients' case data to select the most plausible values for these parameters. This might be achieved in a general Bayesian framework with priors over possible parameter values, or just by determining frequencies in the data.

From the point of view of the medical researcher, the task is to ensure that the causal graph is complete and correct. A common recommendation from AI is to view the links in the causal graph as variable parameters and search for the configuration that is most plausible, given observational and experimental data. Often the search is widened to include minor changes in language, by positing new, unknown common causes in the causal graph to better account for dependencies in the data.

If none of the conjectured networks fits well with background knowledge or further data, then more radical changes in search direction are initiated. A first step might be to acknowledge that a fundamental assumption behind the representation is violated (in this case, the causal Markov condition is the fundamental assumption behind the Bayesian network representation), and to move to representations which do not require this assumption. If the local vicinity of familiar causal representations is exhausted, the task is to articulate new causal representations and extend the search. Devising new causal representations is itself a search problem: we obviously favour simple languages and representations which leave most of our causal intuitions intact, and in a Bayesian search framework one can devise priors which reflect this bias, and select the representation which has maximum posterior probability conditional on the new evidence regarding the failure of known representations.

Under this scheme, abductive problems are all selective: the context of the problem helps to determine the direction of search, and if the search space is viewed as a hierarchy, it is those higher-level transitions, which involve changing key assumptions or language, that are usually considered creative.

In sum, while I agree with Magnani that there is plenty of scope for the field of artificial intelligence to progress our understanding of the logic of abduction, I doubt whether progress will be made through the spurious distinctions touched on here—it is more likely to come from the insights into search that form the core of much AI research.

There is much more to Magnani's book than is encapsulated in its treatment of the key distinctions relating to abduction. For example, chapter 7 discusses an interesting extension of falsificationism which incorporates the concept of negation as failure prevalent in the logic programming literature: the idea is that theories like Freudian psychoanalysis and physics according to Poincaré's conventionalism, which are difficult to falsify, can be rejected on the basis of their failure to yield positive consequences.

Although the book would have greatly benefited from more thorough editing, do not let stylistic considerations put you off. It is to be commended for engaging with ideas from AI, cognitive science, and even medical education, as well as providing a broad guide to much of the literature on abduction.

References

Flach, P. A. and Kakas, A. C. [2000]: 'On the relation between abduction and inductive learning', in Dov M. Gabbay & Rudolf Kruse (*eds*) *Handbook of defeasible reasoning and uncertainty management systems*, Vol. 4: *Abductive reasoning and kearning*, Kluwer Academic Publishers, pp. 1–33.

Galison, P. [1997]: *Image and logic: a material culture of micro-physics*, University of Chicago Press.

Gillies, D. [1996]: *Artificial intelligence and scientific method*, Oxford: Oxford University Press.

Peirce, C. S. [1932–1963]: *Collected papers of C. S. Peirce*, Cambridge, Mass.: Harvard University Press, 1932–1963.

Thagard, P. [1988]: *Computational Philosophy of Science*, Cambridge, Mass.: M.I.T. Press/Bradford Books.