

CHAPTER 9

CONCLUSIONS AND FUTURE WORK

— in which the main contributions of this thesis pass the review, and in passing the main open problems and opportunities for future work are indicated —

§35. PHILOSOPHY

I STARTED MY investigations with an overview and analysis of the philosophy of induction. The by now infamous ‘Problem of Induction’ has a long and confused history in philosophy. Confusions are usually caused by misconceptions, and the case of induction is no exception in this respect. The chief misconception in many philosophical studies of induction is, I believe, the idea that we should be looking for an infallible ‘inductive logic’. A related misconception, which I shall deal with in the next section, is that logic is the study of ‘correct’ reasoning. If we liberate ourselves from these and related misconceptions, we see that it is both possible and desirable to have a precise logical account of induction, without falling into the trap of the ‘deductivist’ approach.

The first step towards a logical account of induction, a step whose significance cannot be overrated, has been made by Charles Sanders Peirce when he distinguished between the process of inductive hypothesis formation on one hand, and the process of evaluating or justifying a chosen hypothesis on the other. The process of forming an inductive hypothesis is a logical process, not in the sense that the hypothesis is obtained by mechanical inference, but in the sense that there are certain necessary conditions imposed upon the logical relation between the hypothesis and the evidence leading to that hypothesis. Those, like the late Sir Karl Popper, who argued against inductive hypothesis formation as being non-logical, were barking up the wrong tree: they wrongly identified logic with the study of reasoning procedures, rather than the study of reasoning forms.

Peirce built his analysis of the logical form of inductive hypothesis formation around the notion of explanation: the hypothesis should be such that it explains the observations, in the sense that the latter are deductive consequences of the former. I have argued that a more comprehensive view of inductive hypothesis formation can be obtained by viewing the logical relation between evidence and hypothesis as a parameter, that can be instantiated to explanation by deductive entailment as proposed by Peirce, but also to, for instance, explanation by plausible entailment, or to quite different logical relations, such as the relation of confirmation studied by Carl G. Hempel.

Even if the statements ‘this evidence is explained by this hypothesis’ and ‘this evidence confirms this hypothesis’ can both give rise to a logic of induction, they have very different logical characteristics. The most significant difference is that explanations

9. *Conclusions and future work*

may be strengthened without ceasing to be an explanation, while confirmed hypotheses may be weakened without being confirmed. Clearly, combining these two opposite characteristics (together with some starting point like ‘the evidence itself is a possible hypothesis’) leads immediately to a situation in which any hypothesis is logically possible given arbitrary evidence. This phenomenon has been characterised as ‘paradoxical’ by many philosophers of science, starting with Hempel — the viewpoint defended in this thesis is that, insofar it is conceived as a problem, it can be solved quite simply by clearly separating the intuitions pertaining to only one of these statements from the intuitions arising only from the other.

Should we, then, take recourse to an extreme relativist position and conclude that the logic of induction is fully dependent upon the viewpoint one prefers? Certainly this would be much too liberal a position, but the conclusion I do draw is that the logical characteristics of inductive hypothesis formation should be studied relative to the goal with which the hypothesis is sought. ‘Peircean’ or explanatory induction, on the one hand, aims at finding explanations of the observations, while ‘Hempelian’ or confirmatory induction, on the other, seeks to find generalisations expressing implicit, non-explanatory regularities displayed by the observations. I will return to this issue in section §37 below.

§36. LOGIC

The boldest deviations of common views voiced in this thesis are probably concerned with the aims and scope of logic. From the perspective put forward in this thesis, logic is much more than the study of correct (deductive) reasoning: it is the study of reasoning forms. As such, it has much more of a *descriptive* than a prescriptive or normative nature, one of its main aims being to provide a catalogue of different reasoning forms and their formalisations. Such a descriptive theory of reasoning forms consists of the following elements:

- (i) one or more logical schemes;
- (ii) instantiations of these schemes, each meant to formalise a certain form of reasoning;
- (iii) proof- and metatheoretical characterisations of these instantiations.

Historically, logic as we know it today has mainly evolved by developing instantiations formalising deductive reasoning in different logical languages (propositional, predicative, modal, temporal, and so forth). The logical scheme that has emerged as a result of these instantiations is the scheme of satisfaction-preserving semantics originally proposed by Tarski. Even if this scheme has developed considerably over the years (*cf.* Kripke’s possible worlds semantics), the message stayed the same: the scheme, and its instantiations, limits attention to non-defeasible forms of reasoning.

The recent developments of so-called ‘non-monotonic’ logics formalising aspects of plausible reasoning indicate that logicians feel an increasing need to liberate themselves from the restrictions imposed by the non-defeasible reasoning scheme. The framework of plausible consequence relations pioneered by Gabbay and Kraus, Lehmann and Magidor indicates that the modifications to the non-defeasible scheme needed to model e.g. preferential reasoning are relatively modest, requiring not more than the addition of a

§36. *Logic*

preference relation over classical models, and defining entailment over the minimal models of the premises under this preference relation, rather than over all models of the premises. Philosophically speaking however, this modest modification represents a drastic departure from the deductive foundations of logic. It is now time, I believe, for logic to start approaching the question from the other end: to reflect upon the essence and nature of reasoning schemes, rather than of some more or less arbitrarily chosen instantiations.

In this thesis I have proposed preservation semantics as a reasoning scheme, which is even more radical in that it strictly subsumes Kraus *et al.*'s preferential semantics. The basic idea is that from the set of semantic objects assigned to the premises of the argument one constructs another set of semantic objects, which need not be included in the first set, such that every one of these is among the semantic objects assigned to the conclusion of the argument. Admittedly, this proposal has not been analysed very thoroughly, since it merely served to reach the right frame of mind for the subsequent logical analysis of inductive hypothesis formation. Nevertheless the proposal has some potential, witnessed by the fact that one can conceive instantiations formalising such different reasoning forms as plausible reasoning, inductive (generality-preserving) reasoning, and counterfactual reasoning. The development of this and similar logical schemes is seen as the major challenge for the field of logic in the years to come.

For deductive logic the semantics offers, besides indicating what semantic quality is preserved, a second function: it estimates the truth of the conclusion of an argument. I have argued that, in the non-deductive case, this second function is performed by a different kind of semantics, *viz.* a truth-estimating semantics. Unlike a preservation semantics, a general truth-estimating semantics, such as Carnap's 'inductive logic', does not give rise to a proof-theory, and is therefore not tied to a specific form of reasoning. Preservation semantics and truth-estimating semantics have complementary roles, only to coincide in the case of deductive reasoning.

Truth-estimating semantics usually give numerical estimates for the probability of the truth of the conclusion, given the truth of the premises, while preservation semantics are commonly couched in qualitative, non-numerical terms. This triggers a fundamental question as to the relation between qualitative and numerical approaches to semantics. Carnap seems to have believed that qualitative approaches like Hempel's relation of confirmation are derivations of the full-fledged numerical approaches, but I doubt that the relation is so straightforward. Both qualitative (i.e. preservatory) and numerical (i.e. truth-estimating) approaches abound in the literature dealing with non-deductive reasoning, and it appears to me that the relation between these two types of semantics establishes a major research area in the field of logic.

§37. ARTIFICIAL INTELLIGENCE

The research question which ignited my investigations was not at all concerned with the philosophical underpinnings of inductive reasoning, nor with the aims and scope of logic. The original research question was to investigate the applicability of approaches to computational induction, as developed in the field of machine learning, to the large collections of data stored in databases. The main problem seemed to be that some of the

9. *Conclusions and future work*

tacit assumptions underlying these approaches didn't quite fit in the database framework, but at the same time it was hard to get a handle on those assumptions.

In retrospect, a lot of those issues have been clarified, not just through my work but also through the work of Luc De Raedt. Both of us started to work on systems inducing integrity constraints from a collection of data. What an integrity constraint is, is perhaps best explained by stating what it is not: a classification rule. What I am stating here in a few words is a fundamental insight that slowly took shape over a number of years, and it was well after that insight had emerged that I discovered almost the same distinction in the philosophical literature. The philosophical and logical elaboration of an issue that is very relevant for the theory and practice of inductive machine learning is what I view as the main contribution of this thesis.

But amidst of all this logico-philosophical weightiness, what are the gems a practical machine learner should bring home? The first message is that there's more to logic than just syntax. Many early 'logical' approaches to machine learning did not employ much more logical apparatus than a logic-based description language. Inductive logic programming acquired the taste of proof procedures, and even added a flavour of (Herbrand) semantics — however, it is still mainly doing logic programming inductively. My thesis can be seen as a contribution to the semantics of programming in inductive logic.

However, as there is hardly any logic program in the whole thesis, its significance extends beyond inductive logic programming to inductive machine learning in general. The concept of a conjectural consequence relation that I introduced in this thesis allows us to reason about different learning situations and to formulate the main characteristics of those learning situations in an unambiguous manner. I have provided a catalogue of such characteristics, and the representation results provide further insights as to how these characteristics interact. Finally, the analysis that has been carried out in this thesis may help us to understand the relation between inductive learning and other approaches to representing and reasoning with knowledge, and hopefully contributes to the development of artificial intelligence into a mature science.

* * * * *