PART II

FOUNDATIONS

The next three, increasingly technical chapters make up the core of this thesis. In the first chapter, Outline of a descriptive theory of induction*, I take issue with the idea that a logical semantics is necessarily based on the notion of truth. In my view, any quality worth preserving in an argument may give rise to a logic formalising some useful form of reasoning. Two of the qualities that may be preserved in inductive arguments are explanatory power and regularity of interpretations. In the next chapter,* Properties of conjectural consequence relations*, I commence my study of general patterns of conjectural reasoning in the spirit of Kraus* et al. *Starting with the adequacy conditions for a material definition of the relation of confirmation formulated by Hempel, I propose a number of rules for explanatory and confirmatory reasoning, and some rules that are meaningful in both cases. Systems of such rules are studied and semantically characterised in the final chapter in this part,* Rule systems for conjectural reasoning.

CHAPTER 5

OUTLINE OF A DESCRIPTIVE THEORY OF INDUCTION

— in which it is argued that the logical foundations of induction be based on a generalised notion of preservation semantics, two qualities that may be preserved in inductive arguments being explanatory power and regularity of interpretations —

IN THIS CHAPTER I will present a blueprint of a descriptive theory of induction: a catalogue of different forms of inductive reasoning. The central concept in such a descriptive theory of induction is the notion of a conjectural consequence relation, formalising the process of inductive hypothesis formation. This concept, which will be fully developed in the next two chapters, is called *conjectural* consequence relation rather than inductive consequence relation, because it may not completely capture every aspect of inductive hypothesis formation (for instance, there is no inherent notion of generalisation), and because it may be applicable to other forms of reasoning as well.

The outline of a descriptive theory of induction presented in this chapter will also discuss other ingredients, and their relation to conjectural consequence relations. In order to get a clear picture of what other ingredients are needed, we start by asking a very general question: What constitutes a logical system? In principle, the answer will be that the usual tripartite model of semantics, proof theory and metatheory will suffice. However, it is clear that such a semantics cannot be satisfaction-preserving, as in the case of the standard semantics for deduction (§14). This raises two questions:

(*i*) are there alternative semantic qualities that can be preserved?
(*ii*) can we say anything about the truth of the conclusion, given the truth of the premises?

The first question will lead to the general notion of a preservation semantics, while the second leads to the concept of a truth-estimating semantics.

On the level of proof theory, I will draw a distinction between a proof procedure and what I call a discovery procedure, which only infers conclusions that maximise some utility measure. Although the latter is, in general, what we want (even in the case of deduction!), I will argue that the utility measure is an extra-logical factor, outside the logical system proper.

5. Outline of a descriptive theory of induction

§18. WHAT IS A LOGICAL SYSTEM?

Recall that in chapter 4 the following picture of logic emerged. Logic can be characterised on three distinct levels: semantics, proof theory, and metatheory. The aim of semantics is to provide formulas with a meaning, and to define the notion of semantic entailment. The goal of proof theory is to provide a more computationally oriented counterpart to semantic entailment, by formulating axioms and inference rules. Semantics and proof theory are linked together by the metatheoretical notions of soundness and completeness.

The question arises whether this tripartite model of semantics, proof theory and metatheory is capable of capturing the essence of an arbitrary form of reasoning. Clearly, when devising a non-deductive logical system we have to give up some of the nice properties of deductive logic. In this light, it is important to have a clear idea of which characteristics of deductive logic are inherent to logical systems as such, and which characteristics derive from the fact that the form of reasoning being modelled is deductive.

This is by no means a simple task, since the standard deductive machinery seems so familiar to us that it is hard to imagine alternatives. This is especially true for the standard deductive semantics, due to Tarski (1936). In the words of John Etchemendy:

> 'The highest compliment that can be paid the author of a piece of conceptual analysis comes not when his suggested definition survives whatever criticism may be leveled against it, or when the analysis is acclaimed unassailable. The highest compliment comes when the suggested definition is no longer seen as the result of conceptual analysis—when the need for analysis is forgotten, and the definition is treated as common knowledge. Tarski's account of the concepts of logical truth and logical consequence has earned him this compliment.' (Etchemendy, 1990, p.1)

Unlike Etchemendy, who criticises it on the grounds of material inadequacy, I accept Tarski's standard deductive semantics, but am seeking ways to generalise it, in order to make it applicable to a wider spectrum of reasoning forms.

What is a semantics?

The distinguishing virtue of Tarski's model-theoretic semantics is that it substantiates the following statement:

> (1) if the premises of a deductive argument are true, then its conclusion is true also.

As we have seen in §14, Tarski achieves this by defining truth as satisfaction by a distinguished intended interpretation. In this way we obtain a clear-cut distinction between deductively valid and deductively invalid arguments.

Clearly, a non-deductive semantics must substantiate a much weaker statement than (1). One possible weakening is embodied in Carnap's proposal for an 'inductive logic' (see §7):

(2) if the premises E of an argument are true, then **the probability that** its conclusion H is true also **is estimated at** $c(H,E)$.

Notice that **every** argument is assigned a degree of confirmation, so that (2), by itself, does not set apart a distinguished set of ‘inductively valid’ arguments. However, we could achieve something like that by stipulating that $c(H,E) > 0$, or $c(H,E) > c(H,\mathbf{true})$, or $c(H,E) = 1$ (the latter stipulation would, presumably, yield the deductively valid arguments).

The KLM framework provides an alternative, qualitative weakening of (1):

(3) if the premises of a plausible argument are true, then its conclusion is **expected to be** true also.

Here, the conclusion of a plausible argument is expected to be true whenever the intended interpretation is possible in at least one most preferred epistemic state among those that satisfy the premises (see Definition 4.6). .

I have tried to demonstrate that different semantic accounts can be formulated in a way which reveals the manner in which they assess the truth of the conclusion, given the truth of the premises. Let us call such an assessment the *truth-estimating conditions* imposed by the semantics. What I want to illustrate now is that the truth-estimating viewpoint is not always the conceptually most appropriate one.

Consider Peirce’s definition of abduction as inference of explanatory hypotheses, where a hypothesis is explanatory if it logically entails the premises of the abductive argument. This leads to the following truth-estimating condition:

(4) if the premises of an abductive argument are false, then its conclusion is false also.

What this truth-estimating condition tells us is that the truth of the premises of an abductive argument is a **necessary** condition for the truth of its conclusion. The weakness of this truth-estimating condition, being inherent to abduction *per se*, is something we have to live with — but its formulation seems to be quite counterintuitive. I will argue below that a much better picture of the semantics of abduction is obtained when it is defined in terms of a semantic quality that is preserved by abductive arguments[54].

Such a preservatory viewpoint has already been considered in §14, where we concluded that deduction is satisfaction-preserving in the sense that an interpretation satisfies the conclusion of a deductive argument whenever it satisfies the premises:

(5) satisfaction by an interpretation is preserved when passing from the premises of a deductive argument to its conclusion.

It is easy to see that (5) is a sufficient condition for (1), if a statement is defined to be true whenever it is satisfied by a distinguished intended interpretation. Conversely, (5) is obtained if (1) holds regardless of which interpretation is the intended one. Thus, (1) and (5) are equivalent if we have no additional knowledge about the intended interpretation.

---

[54] Sometimes, condition (4) is paraphrased by saying that a reasoning form like abduction is falsity-preserving — but who wants to preserve falsity?

In a similar vein, the KLM semantics can be viewed from a preservatory viewpoint:

(6) satisfaction **by a most preferred epistemic state** is preserved when passing from the premises of a plausible argument to its conclusion[55].

The equivalence of (3) and (6) is a matter of definition rather than proof: (6) gives a precise answer to the question what it means to say that the conclusion of a plausible argument is expected to be true. One could say that the truth-estimating viewpoint of KLM semantics is derived from the preservatory viewpoint, which is the more fundamental one.

Abduction or explanatory reasoning is another form of reasoning for which the preservatory viewpoint is much more natural. Consider the following statement:

(7) **explanatory power** is preserved when passing from the premises of an abductive argument to its conclusion.

Explanatory power is a central concept in the explanatory view of induction put forward in this thesis. Intuitively, the explanatory power is the set of explanations of observed phenomena provided by a logical theory. An argument preserves explanatory power if the set of explanations provided by the conclusion is at least as comprehensive as the set of explanations provided by the premises. In the spirit of Hempel's deductive-nomological model of explanations (Hempel, 1966), we may define the explanatory power of a theory as the set of its deductive consequences[56]. Now, one may note that under this definition of explanatory power (7) is equivalent to saying that the premises are among the deductive consequences of the conclusion, which is in turn equivalent to (4). However, it is clear that (7) provides a much more natural viewpoint on the semantics of abduction than (4). Furthermore, there are wider conceptions of explanatory power under which the equivalence of (7) and (4) no longer holds: for instance, we could define the explanatory power of a formula as the set of its plausible, rather than deductive, consequences.

The point I hope to have illustrated is this. There are (at least) two possible angles from which one can view the concept of semantics, one being more natural for some forms of reasoning, the second for others. The preservatory viewpoint was found more natural for explanatory reasoning, and more fundamental for plausible reasoning *à la* KLM. The standard deductive semantics seems to be primarily viewed from the truth-estimating viewpoint, although it can be independently and equivalently viewed from the preservatory viewpoint.

It remains to be decided whether Carnap's degree of confirmation can be equally well

---

[55] This statement is slightly ambiguous. It does **not** mean to say that the quality of 'being most preferred' is preserved (which would imply that every most preferred model of the premises is also a most preferred model of the conclusion), but rather that satisfaction by 'selected models' is preserved (namely, the most preferred models of the premises). This will be made more precise below.

[56] This is, admittedly, a simplification. First of all, it disregards the structure of the proof of the observation from the theory: some explanations may be more powerful, or simple, than others. Secondly, observations may be restricted to a sublanguage of the language in which the theory is formulated. The proposed definition of explanatory power is therefore a starting point, that may require further refinements.

approached from the preservatory angle. My claim is that it cannot — nor is it intended to serve as a preservatory semantics. This becomes immediately clear once one realises that the preservatory viewpoint makes sense only when it indeed serves to dismiss some arguments that do not preserve the designated semantic quality. But Carnap's degree of confirmation is defined for arbitrary arguments — indeed, to restrict its scope would be to limit its applicability unnecessarily. But if we do not restrict the scope of the *c*-function, then the preserved semantic quality must be void.

A related point has been made already by Carnap himself when he remarked that induction, in his view, does not represent inference in the classical sense. Let us recall the citation from chapter 2:

> 'What we call inductive logic is often called the theory of nondemonstrative or nondeductive inference. Since we use the term 'inductive' in the wide sense of 'nondeductive', we might call it the theory of inductive inference... However, it should be noticed that the term 'inference' must here, in inductive logic, not be understood in the same sense as in deductive logic. Deductive and inductive logic are analogous in one respect: both investigate logical relations between sentences; the first studies the relation of [entailment], the second that of degree of confirmation which may be regarded as a numerical measure for a partial [entailment]... The term 'inference' in its customary use implies a transition from given sentences to new sentences or an acquisition of a new sentence on the basis of sentences already possessed. However, only deductive inference is inference in this sense.' (Carnap, 1950, §44B, pp. 205–6)

What is emphasised here is that there is no inherent notion of proof in his 'inductive logic'. According to Carnap, it does not make sense to ask: 'is there an inductive proof of H from E?'; all one can ask is 'what is the degree to which H is confirmed by E?'. What this demonstrates is that *truth-estimating conditions do not necessarily lead to an associated proof theory*.

Let us recover our breath and reinforce the points made above. My first point is that there exist *two different concepts of semantics*, rather than just two viewpoints from which the concept of semantics can be perceived. *Truth-estimating semantics* aims at assessing the truth of the conclusion of an argument. *Preservation semantics* aims at assessing the semantic connection that exists between premises and conclusion of an argument. These two concepts are fundamentally different, only to coincide in one specific case, namely the case of deductive reasoning (fig. 5.1).

My second point is that, in principle, *only the preservation semantics gives rise to a proof theory*. Like the previous point, this point cannot be rigorously proved but only made plausible. One case in point is, of course, provided by Carnap's 'inductive logic', which doesn't have a proof theory (or, alternatively, a trivial one: infer anything from anything). This also means that the concepts of soundness and completeness only make sense in the presence of a preservation semantics.

A third, related point is that, while a preservation semantics mirrors some essential semantic feature of a particular form of reasoning, a truth-estimating semantics is, at least in principle, applicable to arguments of any type. Carnap's claim that his 'inductive

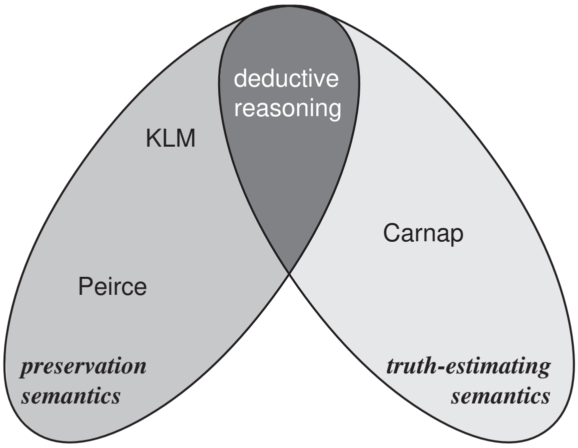

Figure 5.1. The two concepts of preservation semantics and truth-estimating semantics are fundamentally different, but coincide in the case of deductive reasoning.

logic' captures the essence of non-deductive reasoning (and even, as a special case, deductive reasoning) is therefore, in a way, justified. It is not his framework, which definitely has its merits, that is criticised here, but the fact that he calls it a logic. In my view, it embodies a proposal for a truth-estimating semantics that can, in principle, be applied to an arbitrary form of reasoning.

The idea that preservation, rather than mere truth-preservation, is the essential feature of logic is not new. In a paper published in 1991, Jennings, Chan and Dowad 'argue for a generalisation of inference from the standard account in terms of truth preservation to one which countenances preservation of other desirable metalinguistic properties':

> '(…) **truth is not the only inferentially preservable property**. A system of inference essentially provides procedures by which a set Σ of sentences (for example, the set of one's beliefs) having some complex of metalinguistic properties can be unfailingly extended to a larger set Σ′ having the same complex of properties. By all means, we may regard *truth* as one of the properties to be preserved, but what other properties are to be preserved can depend upon our interests.' (Jennings *et al.*, 1991, p.1047)

Furthermore, the view that abduction is the logic of preserving explanations has also been put forward by Zadrozny, who calls an inference rule abductive ‘if it preserves sets of explanations’ (Zadrozny, 1991, p.1).

For the purposes of this thesis, I will now attempt to give a more exact definition of a preservation semantics. The basic idea is that such a semantics specifies a set of semantic objects, such that for an inference from α to β, if a semantic object stands in a certain relation to α (it is for instance a model of α), it stands in the same relation to β; or equivalently, the set of semantic objects $[\![\alpha]\!]$ standing in a certain relation to α is a subset of the set of semantic objects $[\![\beta]\!]$ standing in the same relation to β. A modification of this scheme is suggested by the preferential semantics of Kraus *et al.*: to construct from $[\![\alpha]\!]$ a different set of semantic objects to be included in $[\![\beta]\!]$, by means of a preservation function f.

> DEFINITION 5.1. Let L be a logical language. A *preservation semantics* is a triple $\langle S, [\![\cdot]\!], f\rangle$, where S is a set of *semantic objects*, $[\![\cdot]\!]: L \rightarrow 2^S$ is a *meaning function* mapping any formula to a set of semantic objects, and f: $2^S \rightarrow 2^S$ is a *preservation function* mapping any set of semantic objects to another set of semantic objects. The consequence relation defined by a preservation semantics is given by $\alpha \mid\!\sim \beta$ iff $f([\![\alpha]\!]) \subseteq [\![\beta]\!]$.

As an example, the cumulative KLM semantics is obtained from this scheme by interpreting $[\![\alpha]\!]$ as the set of epistemic states satisfying α, and defining f as the function returning the set of minimal states under a preference ordering. Clearly, the standard deductive semantics is another instance of this scheme, with $[\![\alpha]\!]$ denoting the set of classical models of α, and f the identity function. Finally, to obtain the explanation-preserving semantics of abduction, we interpret $[\![\alpha]\!]$ as the set of consequences of α under some (quasi-)deductive inference mechanism, and f again as the identity function. .

Clearly, it will be necessary to formulate some additional constraints on the functions $[\![\cdot]\!]$ and f. For instance, it is reasonable to require that $[\![\cdot]\!]$ is well-behaved with respect to the logical connectives:

$$[\![\alpha\wedge\beta]\!] = [\![\alpha]\!] \cap [\![\beta]\!]$$

$$[\![\alpha\vee\beta]\!] = [\![\alpha]\!] \cup [\![\beta]\!]$$

$$[\![\neg\alpha]\!] = S - [\![\alpha]\!]$$

and so on. Such constraints will indeed be satisfied by any preservation semantics we will consider in this thesis.

Furthermore, it seems reasonable to assume that f maps any set of semantic objects onto a subset:

$$\forall s\subseteq S\colon f(s) \subseteq s \tag{*}$$

Indeed, this constraint is satisfied for all three preservation semantics just mentioned (in fact, for the standard deductive semantics and the abductive semantics, set-inclusion is trivially strengthened to equality). Nevertheless, I believe that this constraint is too strong for some other forms of reasoning, most notably inductive reasoning and counterfactual reasoning. I will not dive into the subject here — an example will be given below in §19. However, for all the preservation semantics formally worked out in chapter 7 the inclusion constraint (*) holds. In such cases we will say that f is a *selection function*.[57]

With this conception of a preservation semantics we have sufficiently liberated ourselves from the chains imposed by the standard deductive semantics. A complete formal characterisation of this notion, however important, is beyond the scope of this thesis. As for the concept of a truth-estimating semantics, I will not attempt to give a precise definition, since this concept does not play a major role in this thesis. For all practical purposes, it may be identified with a function $c: L \times L \rightarrow [0,1]$, satisfying a number of constraints like in Carnap's 'inductive logic'.

What is a proof procedure?

Having observed the need for a wider conception of the concept of a logical semantics, it is now time to consider the concept of a proof theory. Here, I take the position that the 'traditional' conception of a proof theory is still adequate as a computational counterpart to the preservation semantics.

It follows that metatheory maps the relation between proof theory and preservation semantics. For instance, a proof procedure is sound if any formula it derives preserves the designated semantic quality relative to the premises. Conversely, the proof procedure is complete whenever it is able to derive, from arbitrary premises, every conclusion that preserves that same semantic quality. But now we may ask whether completeness in this sense is always a desirable property. Isn't it true that, even in deductive logic, we are often only interested in deriving only those conclusions that, besides preserving the semantic quality, serve a certain goal?

For instance, suppose that we want to construct an Automated Mathematician: a computer program that, equipped with basic mathematical knowledge, is able to derive interesting mathematical concepts and theorems[58]. Such a computer program should be able, for instance, to derive the fundamental theorem of arithmetic from elementary number theory. Now let Σ denote the initial knowledge, then we may write $\Sigma \vdash_{AM} \varphi$ if φ is a formula derived by the Automated Mathematician. Clearly, we don't want $\vdash_{AM}$ to be complete with respect to the standard deductive semantics[59]: we don't want tautologies or formulas in Σ to be 'derived', nor do we want every logically equivalent formulation of the fundamental theorem of arithmetic to be produced.

On the other hand, if we wanted to implement an Automated Theorem Checker which, supplied with Σ and φ, tests whether $\Sigma \vdash_{ATC} \varphi$, then it is clear that, ideally, $\vdash_{ATC}$ should be complete with respect to $\models$, because the theorem checker should work for arbitrary Σ and φ. Similarly, if $\vdash_{LP}$ denotes a query answering procedure in logic programming systems, constructing a substitution θ such that $\Sigma \vdash_{LP} \varphi\theta$ for arbitrary logic programs Σ

---

[57] Strictly speaking, if f is not a selection function the name of a preservation semantics is a bit of a misnomer, since some of the elements of $f([\![\alpha]\!])$ do not stand in the same relation to α as they stand to β.

[58] See (Lenat, 1980) for an attempt to tackle this problem.

[59] According to Gödel's celebrated incompleteness theorem, it is even impossible for a proof procedure to be complete as soon as elementary number theory is included.

and queries φ, it is again clear that $\vdash_{LP}$ should be complete[60].

One way to describe the behaviour of a *discovery procedure* like the Automated Mathematician is to view it as a heuristic search algorithm which generates, in some way or another, formulas φ such that (*i*) $\Sigma \vdash_{ATC} \varphi$, and (*ii*) $h(\varphi)$ reaches some optimal value, where h is a *heuristic evaluation function* assessing the utility of the formula as a mathematical theorem. There is no need to explicate these concepts any further; suffice it to say that the resulting discovery procedure effectively establishes a restriction of the proof procedure $\vdash_{ATC}$. Clearly, automated induction algorithms embody discovery procedures rather than proof procedures. The question we need to address now is whether the heuristics employed in such discovery procedures are part of the logical system or not.

In this thesis I will take the position that discovery procedures like the Automated Mathematician and automated induction algorithms are built upon logical systems, but are not themselves logical systems. This is not to say that proof procedures are more important than discovery procedures: on the contrary, most practical research on induction in artificial intelligence is aimed at developing inductive discovery procedures rather than inductive proof procedures. Furthermore, it is usually a good idea to integrate the heuristics with the proof procedure, in order to avoid the generation of useless formulas as much as possible. The point I want to stress here is not that discovery procedures are useless, but rather that the heuristics they employ are *extra-logical* and not part of the logical system. Indeed, suppose that we would include heuristics in our logic, i.e. in the proof procedure: in order to characterise the resulting proof procedure, we would need to include them in the semantics as well. What we would have achieved, then, is mainly a corruption of the typical (and meaningful) dualism that exists traditionally between proof theory and semantics. I prefer to keep the dualism, and to view discovery procedures as consisting of a logical system on one hand, and extra-logical factors such as heuristics on the other.

The complete picture

We are now in a position to describe the view of a logical system that I propose in the context of this thesis (fig 5.2). Essentially, it follows the traditional lay-out of a deductive system in that it consists of a semantics, a proof procedure, and a metatheory. The main difference with a deductive logic is that the semantics focuses upon the semantic quality that is preserved by every argument, without necessarily assessing the extent to which a derived conclusion is true. The latter is done by a truth-estimating semantics, such as the one proposed by Carnap under the heading ‘inductive logic’. Furthermore, since such a truth-estimating semantics is, at least in principle, independent of a particular reasoning form, it is only considered to be part of the logical system proper insofar it can be merged with the preservation semantics, as in the case of deductive logic (and, presumably, also in the case of plausible logic, although a truth-estimating semantics is not included in the KLM framework).

---

[60] To be precise: clausal resolution (the proof procedure employed in logic programming systems) is not complete in this sense, because it does not obey the metalevel principle of proof by refutation (see §14). However, it is refutation complete ($\Sigma \cup \{\neg\varphi\}$ is inconsistent whenever $\Sigma \models \varphi$), which is sufficient for a query answering procedure.

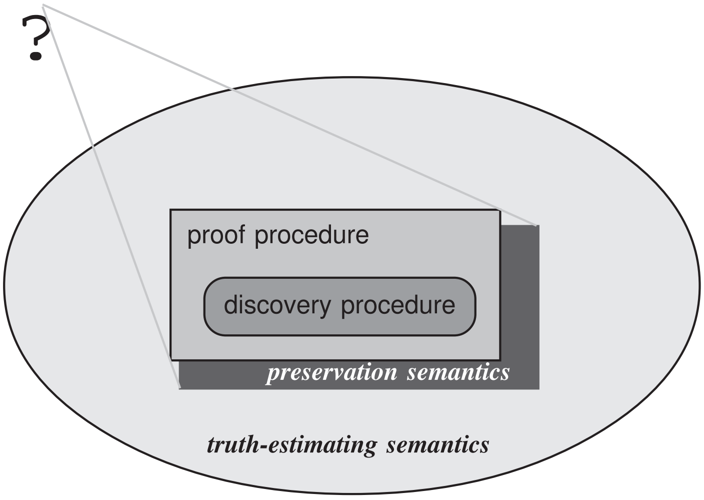

Figure 5.2. A logical system proper consists of a preservation semantics and a proof procedure, sound and complete with respect to the preservation semantics. Truth-estimating semantics, employed to assess the extent to which a derived conclusion is true, is not restricted to the arguments of one particular logical system; a discovery procedure employs heuristics to restrict the range of the proof procedure.

As usual, the metatheory relates the proof procedure to the semantics provided with the logical theory, i.e. the preservation semantics. To state the same point differently: the preservation semantics characterises the proof procedure. Finally, the proof procedure can be extended with heuristics in order to obtain a discovery procedure, which is hoped to derive only the most 'interesting' conclusions. Exactly what constitutes an interesting conclusion is fixed by the heuristics; since such a heuristic is an operational concept that does not have a clear semantic counterpart, I consider it to be an extra-logical factor outside the underlying logical system, even if heuristics are usually operating in close connection with the proof procedure.

§19. ELEMENTS OF AN INDUCTIVE LOGIC

We are now in a position to provide an answer to the question: what constitutes an inductive logic? Basically, this consists of a logical system (semantics, proof procedure, and metatheory), a truth-estimating semantics, and a discovery procedure (i.e., a heuristic utility measure). In this thesis I will concentrate on a logical system for induction, based on notions as explanation and confirmation. I will not further address the problem of devising an appropriate truth-estimating semantics — as I have indicated, I do not think that this problem, important as it may be, is specifically linked to inductive reasoning. The discussion of a discovery procedure for induction will be postponed until §33, where I will describe a utility measure for induced integrity constraints in a relational database.

I will now discuss the two main ingredients for the next two chapters: the concept of a conjectural consequence relation, and several possible preservation semantics for conjectural reasoning. I use the adjective ‘conjectural’ rather than ‘inductive’ because these two ingredients do not exclusively and completely characterise inductive reasoning, both in the sense that some aspects that are typical of induction may be missing (such as the generality of hypotheses, see below), and in the sense that some of the characteristics of induction that are formalised here are shared with certain other forms of reasoning (e.g. the use of a semantic preference ordering).

Conjectural consequence relations

Following the work of Kraus *et al.*, my account of the logical characteristics of induction will be based on the notion of a *conjectural consequence relation*. To clearly mark the distinction with plausible consequence relations, I will borrow a symbol originally proposed by Michalski (1983) and write $\alpha \mathrel{|\!\!\!<} \beta$ for ‘hypothesis β could be conjectured if the evidence is α’.[61] In the next two chapters I will develop several rule systems and semantics for different forms of conjectural reasoning.

We may note again (see also §16) that a consequence relation is a metalevel notion modelling the behaviour of a particular reasoning agent, and that a rule system provides rationality postulates for such an agent. This is no different for conjectural consequence relations. For instance, given background knowledge including

$$\models\texttt{crows\_are\_black}\rightarrow\texttt{chevy\_is\_black}$$

and the fact that the reasoner accepts the inductive argument

$$\texttt{chad\_is\_black} \mathrel{|\!\!\!<} \texttt{crows\_are\_black}$$

our rationality postulates might prescribe that the hypothesis ‘all crows are black’ can still be maintained after observation of Chevy being black, by stipulating that the inductive argument

$$\texttt{chad\_is\_black}\land\texttt{chevy\_is\_black} \mathrel{|\!\!\!<} \texttt{crows\_are\_black}$$

must also be in the consequence relation describing the behaviour of the inductive agent. We could say that a system for conjectural reasoning puts limits to the rational behaviour of an inductive reasoner.

Preservation semantics

The main concern of this thesis is to investigate logical systems for induction. Within such a logical system, preservation semantics occupies a central position. I will now

---

[61] Michalski interprets $\alpha \mathrel{|\!\!\!<} \beta$ as ‘α generalises to β’ (Michalski, 1983, p.89).

investigate a few specific proposals for such a preservation semantics, some of which are worked out in the next chapters.

For explanatory induction the basic idea is to consider explanatory power as the preserved quality, where explanatory power is identified with the set of formulas that can be inferred under a deductive or quasi-deductive regime. Somewhat more precisely, a preservation semantics for explanatory reasoning is built upon an *explanation mechanism*, which in this thesis will be identified with a consequence relation satisfying one of the KLM rule systems **C**, **P** or **M**. Given an explanation mechanism $\mathrel{|\!\sim}$, the meaning function $[\![\cdot]\!]$ maps a formula α to its *closure* under $\mathrel{|\!\sim}$, denoted by $Cn_{\mathrel{|\!\sim}}(\alpha) = \{\gamma \mid \alpha \mathrel{|\!\sim} \gamma\}$[62]. The selection function f is the identity function — that is, an argument $\alpha \mathrel{|\!<} \beta$ preserves explanatory power if $Cn_{\mathrel{|\!\sim}}(\alpha) \subseteq Cn_{\mathrel{|\!\sim}}(\beta)$, i.e. for every γ, if $\alpha \mathrel{|\!\sim} \gamma$ then $\beta \mathrel{|\!\sim} \gamma$.

Clearly, if $\mathrel{|\!\sim}$ is a deductive proof procedure, then this can be reduced to the requirement that $\beta \mathrel{|\!\sim} \alpha$. More precisely, we have the following lemma.

> LEMMA 5.2. *If $\mathrel{|\!\sim}$ is transitive and reflexive, then $Cn_{\mathrel{|\!\sim}}(\alpha) \subseteq Cn_{\mathrel{|\!\sim}}(\beta)$ iff $\beta \mathrel{|\!\sim} \alpha$.*
> *Proof*. For the if part, suppose $\beta \mathrel{|\!\sim} \alpha$ and $\alpha \mathrel{|\!\sim} \gamma$, then $\beta \mathrel{|\!\sim} \gamma$ by Transitivity.
> For the only-if part, suppose $Cn_{\mathrel{|\!\sim}}(\alpha) \subseteq Cn_{\mathrel{|\!\sim}}(\beta)$, then (since $\alpha \in Cn_{\mathrel{|\!\sim}}(\alpha)$ by Reflexivity) $\alpha \in Cn_{\mathrel{|\!\sim}}(\beta)$.

constructed out of the models of α by means of the preservation function f. I will call such a form of reasoning *regularity-based* confirmatory reasoning.

What this notion of regular interpretations brings us is an associated notion of generalisation: the confirmed hypothesis should be a formula that does not refer to the objects in the observations, hence a formula that does more than just replicating the logical content of the observations. We could say that this notion of regularity as invariance under renaming of objects amounts to a view of induction as **generality-preserving** reasoning. This appears to be a very appealing view indeed, which deserves to be investigated further. However, it seems to me that this view has some quite drastic logical ramifications. For one thing, note that f is not a selection function ($f(s)\not\subseteq s$ for some $s\subseteq S$), since some regular interpretations do not satisfy the observations. As a result, quite a few of Hempel's adequacy conditions would fail to hold, such as the entailment condition (C1) and the consequence condition (C2).

For this reason, the regularity-based confirmatory semantics will not be analysed to its full strength in this thesis. Instead we will consider a more 'classical' instantiation in which f is a simple selection function. One possible selection function stays close to the ideas of Hempel, Helft and De Raedt, by treating only the truth-minimal model(s) of the observations as regular. The semantic quality that is preserved is then satisfaction by the regular model. The idea of minimising truth is closely related to the Closed-World Assumption as it is frequently used for reasoning about the information contained in databases; it will be referred to as *closed-world reasoning*. Since the truth-minimal models can be viewed as the minimal elements under a certain partial ordering, closed-world reasoning establishes a kind of preferential reasoning. The general case of an arbitrary preference ordering, which stays very close to the KLM system **P**, is axiomatised in §27.

A third type of conjectural reasoning that will be studied in chapter 7 is similar to closed-world reasoning in that it also employs a partial ordering on semantic objects, namely the information ordering on partial interpretations. However, the pragmatics of this ordering is quite different from a preference ordering: its minimal elements do not represent the 'best' interpretations, but the ones that can be used to enumerate all the others. In other words, the information ordering is a computational rather than a logical device. I will demonstrate in chapter 7 that this form of reasoning is rather weak, requiring no more than compatibility of evidence and inductive conclusion. This form of reasoning is therefore called *consistency-based confirmatory reasoning*.

§20. SUMMARY AND CONCLUSIONS

In this chapter, we set out to present a blueprint of a descriptive theory of inductive reasoning. I have identified conjectural consequence relations as the central notion in such a descriptive theory, modelling certain aspects of the process of inductive hypothesis formation. I have argued that, just like other abstract forms of reasoning, conjectural consequence relations can be formalised by means of a tripartite logical system consisting of semantics, proof theory and metatheory. Different forms of inductive reasoning correspond to different logical systems for conjectural consequence relations, that will be developed in the next two chapters.

It is important to understand the significance of such logical systems for induction. In a way, they accomplish the same as any logical system: a formalisation of the notion of a valid argument. However, validity should here not be understood as deductive validity, i.e. truth preservation: clearly, a deductive logic of induction is impossible. The notion of validity is not absolute, but relative to a certain reasoning form, formalised by a logical system. I believe that much of the confusion about the infamous ‘Problem of Induction’ has been caused by the misconception that a logic of induction should be built on top of the deductive notion of validity.

Much of the work in this chapter has therefore been devoted to a liberation from the dogma that logic is the study of deductive reasoning. It may seem that we have completely overthrown the conventional view of logic — however, the only significant innovation has been to disconnect the notions of preservation semantics (the counterpart of a proof procedure) and truth-estimating semantics. I have also pointed out a distinction between a proof procedure and a discovery procedure, but this distinction is significant for any logic and, although not always articulated so explicitly, probably well understood.

An analysis of some possible logical systems for inductive hypothesis formation will be carried out in the remainder of this thesis. However, this analysis should be viewed as a starting point of the investigations into this subject, rather than an endpoint. One open problem that has been identified is the formalisation of induction as generalisation-preserving reasoning. This idea will not be further pursued here because it would represent, in the current context, too radical a departure from the ideas of Peirce and Hempel.

* * * * *