

PETER A. FLACH

LOGICAL CHARACTERISATIONS OF INDUCTIVE LEARNING

ABSTRACT: This chapter presents a logical analysis of induction. Contrary to common approaches to inductive logic that treat inductive validity as a real-valued generalisation of deductive validity, we argue that the only logical step in induction lies in hypothesis *generation* rather than evaluation. Inspired by the seminal paper of Kraus, Lehmann and Magidor [Kraus *et al.*, 1990] we analyse the logic of inductive hypothesis generation on the meta-level of consequence relations. Two main forms of induction are considered: explanatory induction, aimed at inducing a general theory explaining given observations, and confirmatory induction, aimed at characterising completely or partly observed models. Several sets of meta-theoretical properties of inductive consequence relations are considered, each of them characterised by a suitable semantics. The approach followed in this chapter is extensively motivated by referring to recent and older work in philosophy, logic, and machine learning.

1 INTRODUCTION

This chapter aims to develop a logical account of inductive reasoning, one of the most important ways to synthesize new knowledge. Induction provides an idealized model for empirical sciences, where one aims to develop general theories that account for phenomena observed in controlled experiments. It also provides an idealized model for cognitive processes such as learning concepts from instances. The advent of the computer has suggested new inductive tasks such as program synthesis from examples of input-output behaviour and knowledge discovery in databases, and the application of inductive methods to artificial intelligence problems is an active research area, which has displayed considerable progress over the last decades.

On the foundational side, however, our understanding of the essentials of inductive reasoning is fragmentary and confused. Induction is usually defined as inference of general rules from particular observations, but this slogan can hardly count as a definition. Clearly some rules are better than others for given observations, while yet other rules are totally unacceptable. A logical account of induction should shed more light on the relation between observations and hypotheses, much like deductive logic formalises the relation between theories and their deductive consequences.

This is by no means an easy task, and anyone claiming to provide a definitive solution should be approached sceptically. This chapter suggests a novel perspective by combining older work in philosophy of science with a methodology suggested by recent work in formalising nonmonotonic reasoning. This perspective provides us with a *descriptive* — rather than *prescriptive* — account of induction, which clearly indicates both the opportunities for and limitations of logical analysis when it comes to modelling induction.

We will take it for granted that there exists a distinct and useful form of reasoning called induction. As a model for this form of reasoning we may take

the approaches to learning classification rules from examples that can be found in the machine learning literature, or the work on inducing Prolog programs and first-order logical theories from examples in the recently established discipline of inductive logic programming. By taking this position we will avoid the controversies abounding in philosophy of science as to whether or not science proceeds by inductive methods. This is not to say that philosophical considerations will be completely ignored — in fact, my approach has been partly motivated by works from the philosophers Charles Sanders Peirce and Carl G. Hempel, as I will explain shortly.

The main question addressed in this chapter is the following: *Can we develop a logical account of induction that is sufficiently similar to the modern account of deduction?* By ‘the modern account of deduction’ I mean the approach, developed in the first half of this century and standard by now, of defining a logical language, a semantical notion of deductive consequence, and a proof system of axioms and inference rules operationalising the relation of deductive consequence. By the stipulation that the logical account of induction be ‘sufficiently similar’ to this modern account of deduction I mean that the former should likewise consist of a semantical notion of *inductive consequence*, and a corresponding proof system.

Those perceiving logic as the ‘science of correct reasoning’ will now object that what I am after is a deductive account of induction, and it is known already since Hume that inductive hypotheses are necessarily defeasible. My reply to this objection is that it derives from a too narrow conception of logic. In my view, *logic is the science of reasoning*, and it is the logician’s task to develop formal models of every form of reasoning that can be meaningfully distinguished. In developing such formal models for nondeductive reasoning forms, we should keep in mind that deduction is a highly idealized and restricted reasoning form, and that we must be prepared to give up some of the features of deductive logic if we want to model reasoning forms that are less perfect, such as induction.

The fundamental question then is: which features are inherent to logic *per se*, and which are accidental to deductive logic? To illustrate this point, consider the notion of truth-preservation: whenever the premisses are true, the conclusion is true also. It is clear that truth-preservation must be given up as soon as we step out of the deductive realm. The question then arises whether a logical semantics is mainly a tool for assessing the truth of the conclusion given the truth of the premisses, or whether its main function is rather to define what property is preserved when passing from premisses to conclusion. We will address this and similar fundamental questions in this chapter.

Another objection against the approach I propose could be that deductive logic is inherently prescriptive: it clearly demarcates the logical consequences one *should* accept on the basis of given premisses, from the ones one *should not* accept. Clearly, our understanding of induction is much too limited to be able to give a prescriptive account of induction. My reply to this objection is that, while such a demarcation is inherent to logic, its interpretation can be either prescriptive or descriptive. The inductive logics I propose distinguish between hypotheses one

should not accept on the basis of given evidence, relative to a certain goal one wants the hypothesis to fulfill, and hypothesis one *might* accept. Put differently, these inductive logics formalise the logic of inductive hypothesis generation rather than hypothesis selection, which I think is the best one can hope to achieve by purely logical means.

The objective pursued in this chapter, then, is to develop semantics and proof systems for inductive hypothesis generation. What is new here is not so much this objective, which has been pursued before (see e.g. [Delgrande, 1987]), but the meta-theoretical viewpoint taken in this chapter, which I think greatly benefits our understanding of the main issues. This meta-theoretical viewpoint has been inspired by the seminal paper of Kraus, Lehmann and Magidor [Kraus *et al.*, 1990], where it is employed to unravel the fundamental properties of nonmonotonic reasoning. Readers familiar with the paper of Kraus *et al.* may alternatively view the present chapter as a constructive proof of the thesis that their techniques in fact establish a *methodology*, by demonstrating how these techniques can be successfully applied to analyse a rather different form of reasoning.

The chapter is structured as follows. In Section 2 the philosophical, logical, and machine learning backgrounds of this chapter are surveyed. Section 3 introduces the main logical tool employed in this chapter: the notion of a meta-level consequence relation. Sections 4 and 5 form the technical core of this chapter, stating representation theorems characterising sets of meta-level properties of explanatory induction and confirmatory induction, respectively. In Section 6 we discuss the implications of the approach taken and results obtained in this chapter. Section 7 reviews the main conclusions in a wider context.

2 BACKGROUNDS

This section reviews a number of related approaches from the philosophical, logical, and machine learning literature. With such a complex phenomenon as induction, one cannot hope to give an overview that can be called complete in any sense — I will restrict attention to those approaches that either can be seen as precursors to my approach, or else are considered as potential answers to my objectives but rejected upon closer inspection. We start with the latter.

2.1 *Inductive probability*

By now it is commonplace to draw a connection between inductive reasoning and probability calculus. Inductive or subjective probability assesses the degree to which an inductive agent is willing to accept a hypothesis on the basis of available evidence. A so-called posterior probability of the hypothesis after observing the evidence is obtained by applying Bayes' theorem to the probability of the hypothesis prior to observation. Rudolf Carnap has advocated the view that inductive probability gives rise to a system of inductive logic [Carnap, 1950]. Briefly,

Carnap defines a function $c(H, E)$ assigning a *degree of confirmation* (a number between 0 and 1) to a hypothesis H on the basis of evidence E . This function generalises the classical notion of logical entailment — which can be seen as a ‘confirmation function’ from premisses and conclusion to $\{0, 1\}$ — to an inductive notion of ‘partial entailment’:

‘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]

This citation succinctly summarises why inductive probability is *not* suitable, in my view, as the cornerstone of a logic of induction. My two main objections are the following.

Inductive probability treats all nondeductive reasoning as inductive. This runs counter to one of the main assumptions of this chapter, namely that induction is a reasoning form in its own right, which we want to characterise in terms of properties it enjoys rather than properties it lacks. A more practical objection is that a single logical foundation for all possible forms of nondeductive reasoning is likely to be rather weak. Indeed, I would argue that in many forms of reasoning the goal that is to be fulfilled by the hypothesis, such as explaining the observations, is not reducible to a degree of confirmation.¹

Inductive probability, taken as partial entailment, leads to a degenerated view of logic. This is essentially what Carnap notes when he states that his inductive logic does not establish inference in the same sense as deductive logic (although he would not call it a degeneration). This means that, for instance, the notion of a proof reduces to a calculation of the corresponding degree of confirmation. A possible remedy is to define and axiomatise a qualitative relation of confirmation, such as the relation defined by $qc(H, E) \Leftrightarrow c(H, E) > c(H, \mathbf{true})$. However, such a qualitative relation of confirmation can also be postulated without reference to numerical degrees of confirmation, which would give us much more freedom to investigate the relative merits of different axiom systems. In fact, this is the course of action taken by Hempel, as we will see in the next section.

¹Note that degree of confirmation, in the sense defined by Carnap, is not a quantity that is simply to be maximised, since this would lead us straight back into deductive logic.

I should like to stress that it is not inductive probability or Bayesian belief measures as such which are criticised here — on the contrary, I believe these to be significant approaches to the important problem of how to update an agent's beliefs in the light of new information. Since belief measures express the agent's subjective estimates of the truth of hypotheses, let us say that inductive probability and related approaches establish a *truth-estimating procedure*. My main point is that such truth-estimating procedures are, generally speaking, *complementary* to logical systems. Truth-estimating procedures answer a type of question which nondeductive logical systems, in general, cannot answer, namely: how plausible is this hypothesis given this evidence? The fact that deductive logic incorporates such a truth-estimating procedure is accidental to deductive reasoning; the farther one moves away from deduction, the less the logical system has to do with truth-estimation. For instance, the gap between logical systems for nonmonotonic reasoning and truth-estimating procedures is much smaller than the gap between the latter and logical systems for induction. Indeed, one may employ one and the same truth-estimating procedure for very different forms of reasoning.

2.2 *Confirmation as a qualitative relation*

Carl G. Hempel [1943; 1945] developed a qualitative account of induction that will form the basis of the logical system for what I call confirmatory induction (Section 5). Carnap rejected Hempel's approach, because he considered a quantitative account of confirmation as more fundamental than a qualitative account. However, as explained above I think that the two are conceived for different purposes: a function measuring degrees of confirmation can be used as a truth-estimating procedure, while a qualitative relation of confirmation can be used as the cornerstone for a logical system. I also consider the two as relatively independent: a qualitative confirmation relation that cannot be obtained from a numerical confirmation function is not necessarily ill-conceived, as long as the axioms defining the qualitative relation are intuitively meaningful.

Hempel's objective is to develop a material definition of confirmation. Before doing so he lists a number of adequacy conditions any such definition should satisfy. Such adequacy conditions can be seen as meta-level axioms, and we will discuss them at some length. The following conditions can be found in [Hempel, 1945, pp. 103–106, 110]; logical consequences of some of the conditions are also stated.

- (H1) *Entailment condition*: any sentence which is entailed by an observation report is confirmed by it.
- (H2) *Consequence condition*: if an observation report confirms every one of a class K of sentences, then it also confirms any sentence which is a logical consequence of K .

- (H2.1) *Special consequence condition*: if an observation report confirms a hypothesis H , then it also confirms every consequence of H .
- (H2.2) *Equivalence condition*: if an observation report confirms a hypothesis H , then it also confirms every hypothesis which is logically equivalent with H .
- (H2.3) *Conjunction condition*: if an observation report confirms each of two hypotheses, then it also confirms their conjunction.
- (H3) *Consistency condition*: every logically consistent observation report is logically compatible with the class of all the hypotheses which it confirms.
 - (H3.1) Unless an observation report is self-contradictory, it does not confirm any hypothesis with which it is not logically compatible.
 - (H3.2) Unless an observation report is self-contradictory, it does not confirm any hypotheses which contradict each other.
- (H4) *Equivalence condition for observations*: if an observation report B confirms a hypothesis H , then any observation report logically equivalent with B also confirms H .

The entailment condition (H1) simply means that entailment ‘might be referred to as the special case of *conclusive* confirmation’ [Hempel, 1945, p. 107]. The consequence conditions (H2) and (H2.1) state that the relation of confirmation is closed under weakening of the hypothesis or set of hypotheses (H_1 is weaker than H_2 iff it is logically entailed by the latter). Hempel justifies this condition as follows [Hempel, 1945, p. 103]: ‘an observation report which confirms certain hypotheses would invariably be qualified as confirming any consequence of those hypotheses. Indeed: any such consequence is but an assertion of all or part of the combined content of the original hypotheses and has therefore to be regarded as confirmed by any evidence which confirms the original hypotheses.’ Now, this may be reasonable for single hypotheses (H2.1), but much less so for sets of hypotheses, each of which is confirmed separately. The culprit can be identified as (H2.3), which together with (H2.1) implies (H2). A similar point can be made as regards the consistency condition (H3), about which Hempel remarks that it ‘will perhaps be felt to embody a too severe restriction’. (H3.1), on the other hand, seems to be reasonable enough; however, combined with the conjunction condition (H2.3) it implies (H3).

We thus see that Hempel’s adequacy conditions are intuitively justifiable, except for the conjunction condition (H2.3) and, *a fortiori*, the general consequence condition (H2). On the other hand, the conjunction condition can be justified by a completeness assumption on the evidence, as will be further discussed in Section 5. We close this section by noting that Hempel’s material definition of the relation of confirmation of a hypothesis by evidence roughly corresponds to what we would nowadays call ‘truth of the hypothesis in the truth-minimal Herbrand

model of the evidence’ (see [Lachiche, 2000] for a discussion of closed-world assumptions in induction and abduction). We will return to material definitions of qualitative confirmation in Section 2.4.

2.3 Confirmation vs. explanation

Preceding Hempel’s work on confirmation by almost half a century is the work of Charles Sanders Peirce on abduction: the process of forming explanatory hypotheses. For a full account of Peirce’s theories of abduction and induction the reader is referred to the introductory chapter to this volume; here, a short summary suffices.²

In Peirce’s view, inductive reasoning consists of two steps: (i) formulating a conjecture, and (ii) evaluating the conjecture. Both steps take the available evidence into account, but in quite different ways and with different goals. The first step requires that the conjectured hypothesis *explains* the observations; having a definite logical form, it represents a form of inference. The second step evaluates how well predictions offered by the hypothesis agree with reality; it is not inference, but assigns a numerical value to a hypothesis. The first step corresponds to what I call explanatory hypothesis generation, and the second to hypothesis evaluation.

Leaving a few details aside, Peirce’s definition of explanatory hypothesis generation can be formalised as the inference rule

$$\frac{C, A \vDash C}{A}$$

I propose to generalise Peirce’s definition by including the relation of ‘is explained by’ as a parameter. This is achieved by lifting the explanatory inference from C to A to the meta-level, as follows:

$$\frac{A \vDash C}{C \vDash A}$$

The symbol \vDash stands for the explanatory consequence relation. Axiom systems for this relation will be considered in Section 4.

We have now encountered two fundamental notions that play a role in inductive hypothesis generation: one is that the hypothesis should be *confirmed* by the evidence (Hempel), the other that the hypothesis should *explain* the evidence (Peirce). Couldn’t we try and build the requirement that the hypothesis be explanatory into our definition of confirmed hypothesis?

The problem is that an unthoughtful combination of explanation and confirmation can easily lead into paradox. Let H_1 and H_2 be two theories such that the latter includes the former, in the sense that everything entailed by H_1 is also entailed by

²Here, we are mainly interested in Peirce’s later, inferential theory. See [Flach and Kakas, 2000] for a collection of papers on the relation between induction and abduction.

H_2 . Suppose E is confirming evidence for H_1 ; shouldn't we conclude that it confirms H_2 as well? To borrow an example of Hempel: 'Is it not true, for example, that those experimental findings which confirm Galileo's law, or Kepler's laws, are considered also as confirming Newton's law of gravitation?' [Hempel, 1945, p. 103]. This intuition is formalised by the following condition:

(H5) *Converse consequence condition*: if an observation report confirms a hypothesis H , then it also confirms every formula logically entailing H .

The problem is, however, that this rule is incompatible with the special consequence condition (H2.1). This can be seen as follows: in order to demonstrate that E confirms H for arbitrary E and H , we note that E confirms E by (H1), so by the converse consequence condition E confirms $E \wedge H$; but then E confirms H by (H2.1). Thus, we see that the combination of two intuitively acceptable conditions leads to a collapse of the system into triviality, a clearly paradoxical situation.

Hempel concludes that one cannot have both (H2.1) and (H5), and drops the latter. His justification of this decision is however unconvincing, which is not surprising since neither is *a priori* better than the other: *they formalise different intuitions*. While (H2.1) formalises a reasonable intuition about confirmation, (H5) formalises an equally reasonable intuition about *explanation*:

(H5') if an observation report is explained by a hypothesis H , then it is also explained by every formula logically entailing H .

In this chapter I defend the position that Hempel was reluctant to take, namely that with respect to inductive or scientific hypothesis generation there is more than one possible primitive notion: the relation 'confirms' between evidence and hypothesis, and the relation 'is explained by'. Each of these primitives gives rise to a specific form of induction. This position is backed up by recent work in machine learning, to which we will turn now.

2.4 Inductive machine learning

Without doubt, the most frequently studied induction problem in machine learning is concept learning from examples. Here, the observations take the form of descriptions of instances (positive examples) and non-instances (negative examples) of an unknown concept, and the goal is to find a definition of the concept that correctly discriminates between instances and non-instances. Notice that this problem statement is much more concrete than the general description of induction as inference from the particular to the universal: once the languages for describing instances and concepts are fixed, the desired relation between evidence and hypothesis is determined. A natural choice is to employ a predicate for the concept to be learned, and to use constants to refer to instances and non-instances. In this way, a classification of an instance can be represented by a truth value, which can be obtained by setting up a proof. We then obtain the following general problem definition:

Problem: Concept learning from examples in predicate logic.

- Given:*
- (1) A predicate-logical language.
 - (2) A predicate representing the target concept.
 - (3) Two sets P and N of ground literals of this predicate, representing the positive and negative examples.
 - (4) A background theory T containing descriptions of instances.

Determine: A hypothesis H within the provided language such that

- (i) for all $p \in P : T \cup H \models p$;
- (ii) for all $n \in N : T \cup H \not\models n$.

Notice that condition (ii) is formulated in such a way that the hypothesis only needs to contain *sufficient conditions* for concept membership (since a negative classification is obtained by negation as failure). This suggests an analogy between concept definitions and Horn clauses, which can be articulated by allowing (possibly recursive) logic programs as hypotheses and background knowledge, leading us into the field of inductive logic programming (ILP) [Muggleton and De Raedt, 1994]. Furthermore, P and N may contain more complicated formulae than ground facts. The general problem statement then becomes: given a partial logic program T , extend it with clauses H such that every formula in P is entailed and none of the formulae in N .

The potential for inductive methods in artificial intelligence is however not exhausted by classification-oriented approaches. Indeed, it seems fair to say that most knowledge implicitly represented by extensional databases is non-classificatory. Several researchers have begun to investigate non-classificatory approaches to knowledge discovery in databases. For instance, in previous work I have demonstrated that the problem of inferring the set of functional and multivalued attribute dependencies satisfied by a database relation can be formulated as an induction problem [Flach, 1990; Flach, 1993; Flach, 1995; Flach and Savnik, 1999]. Furthermore, De Raedt and Bruynooghe have generalised the classificatory ILP-setting in order to induce non-Horn clauses from ground facts [De Raedt and Bruynooghe, 1993]. Both approaches essentially employ the following problem statement.

Problem: Knowledge discovery

- Given:*
- (1) A predicate-logical language.
 - (2) Evidence E .

Determine: A hypothesis H within the provided language such that

- (i) H is true in a model m_0 constructed from E ;
- (ii) for all g within the language, if g is true in m_0 then $H \models g$.

Essentially, the model m_0 employed in the approaches by De Raedt and Bruynooghe and myself is the truth-minimal Herbrand model of the evidence.³ The hypothesis is then an axiomatisation of all the statements true in this model, including non-classificatory statements like ‘everybody is male or female’ and ‘nobody is both a father and a mother’.

The relation between the classificatory and non-classificatory approaches to induction is that they both aim at extracting similarities from examples. The classificatory approach to induction achieves this by constructing a single theory that entails all the examples. In contrast, the non-classificatory approach achieves this by treating the examples as a model — a description of the world that may be considered *complete*, at least for the purposes of constructing inductive hypotheses. The approach is justified by the assumption that *the evidence expresses all there is to know about the individuals in the domain*. Such a completeness assumption is reminiscent of the Closed World Assumption familiar from deductive databases, logic programming, and default reasoning — however, in the case of induction its underlying intuition is quite different. As Nicolas Helft, one of the pioneers of the non-classificatory approach, puts it:

‘induction assumes that the similarities between the observed data are representative of the rules governing them (...). This assumption is like the one underlying default reasoning in that a priority is given to the information present in the database. In both cases, some form of “closing-off” the world is needed. However, there is a difference between these: loosely speaking, while in default reasoning the assumption is “what you are not told is false”, in similarity-based induction it is “what you are not told looks like what you are told”.’ [Helft, 1989, p. 149]

There is a direct connection between Peirce’s conception of abduction as generation of explanatory hypotheses and the classificatory induction setting, if one is willing to view a theory that correctly classifies the examples as an *explanation* of those examples. I suggest to draw a similar connection between Hempel’s conception of confirmation as a relation between evidence and potential hypotheses and the non-classificatory induction setting outlined above. Non-classificatory induction aims at constructing hypotheses that are *confirmed* by the evidence, without necessarily explaining them. Rather than studying material definitions of what it means to explain or be confirmed by evidence, as is done in the works referred to above, in the following sections I will be concerned with the logical analysis of the abstract notions of explanation and confirmation.

³An alternative approach is to consider the information-minimal partial model of the evidence [Flach, 1995].

3 INDUCTIVE CONSEQUENCE RELATIONS

In the sections to follow I employ the notion of a consequence relation, originating from Tarski [1994] and further elaborated by Gabbay [1985], Makinson [1989], and Kraus, Lehmann and Magidor [1990; 1992]. In this section I give an introduction to this important meta-logical tool that is largely self-contained. The basic definitions are given in Section 3.1. In Section 3.2 I consider some general properties of consequence relations that arise when modelling the reasoning behaviour of inductive agents. Section 3.3 is devoted to a number of considerations regarding the pragmatics of consequence relations in general, and inductive consequence relations as used in this chapter in particular.

3.1 Consequence relations

We distinguish between the language L in which an inductive agent formulates premisses and conclusions of inductive arguments, and the meta-language in which statements about the reasoning behaviour of the inductive agent are expressed. Throughout the chapter L is a propositional language⁴ over a fixed countable set of proposition symbols, closed under the usual logical connectives. We assume a set of propositional models U , and a satisfaction relation $\models \subseteq U \times L$ that is well-behaved with respect to the logical connectives and compact. As usual, we write $\models \alpha$ for $\forall m \in U : m \models \alpha$, for arbitrary $\alpha \in L$. Note that U may be a proper subset of the set of all truth-assignments to proposition symbols in L , which would reflect prior knowledge or *background knowledge* of the inductive agent. Equivalently, we may think of U as the set of models of an implicit background theory T , and let $\models \alpha$ stand for ‘ α is a logical consequence of T ’.

The meta-language is a restricted predicate language built up from a unary meta-predicate \models in prefix notation (standing for validity with respect to U in L) and a binary meta-predicate \llcorner in infix notation (standing for inductive consequence). In referring to object-level formulae from L we employ a countable set of meta-variables $\alpha, \beta, \gamma, \delta, \dots$, the logical connectives from L (acting like function symbols on the meta-level), and the meta-constants **true** and **false**. Meta-level literals are atomic formulae or their negation; instead of $\neg(\models \alpha)$ we write $\not\models \alpha$, and instead of $\neg(\alpha \llcorner \beta)$ we write $\alpha \not\llcorner \beta$. Formulae of the meta-language, usually referred to as *rules* or *properties*, are of the form $P_1, \dots, P_n / Q$ for $n \geq 0$, where P_1, \dots, P_n and Q are literals (atomic formulae or their negation). Intuitively, such a rule should be interpreted as an implication with *antecedent* P_1, \dots, P_n (interpreted conjunctively) and *consequent* Q , in which all variables are implicitly universally quantified. An example of such a rule, written in an expanded Gentzen-

⁴One may argue that in induction the distinction between statements about individuals (ground facts) and statements about sets of individuals (universal sentences) is crucial, which calls for a predicate language. This point is discussed in Section 6.

style notation, is

$$\frac{\vDash \alpha \wedge \beta \rightarrow \gamma \quad , \quad \alpha \vDash \beta}{\alpha \wedge \neg \gamma \not\vDash \beta}$$

This is a rule with two positive literals in its antecedent, and a negative literal in its consequent. Intuitively, it expresses that an inductive hypothesis β , previously inferred from evidence α , should be withdrawn if the negation of a consequence of α and β together is added to the evidence. A *rule system* is a set of such meta-level rules, denoted by abbreviations in boldface capitals.

Consequence relations provide the semantics for this meta-language, by fixing the meaning of the meta-predicate \vDash . Formally, a consequence relation is a subset of $L \times L$. They will be used to model part or all of the reasoning behaviour of a particular reasoning agent, by listing a number of *arguments* (pairs of *premiss* and *conclusion*) the agent is prepared to accept.⁵ A consequence relation satisfies a rule whenever it satisfies all instances of the rule, and violates it otherwise, where an instance of a rule is obtained by replacing the variables of the rule with formulae from L . A consequence relation satisfies an instance of a rule if, whenever it satisfies the literals in the antecedent of the rule, it also satisfies the consequent. Finally:

- a ground literal $\vDash \alpha$ is satisfied whenever the propositional formula from L denoted by α is true in every model in U ;
- a ground literal $\alpha \vDash \beta$ is satisfied whenever the pair of propositional formulae from L denoted by α and β is an element of the consequence relation.

For instance, the consequence relation $\{\langle p, q \rangle, \langle p \wedge \neg p, q \rangle\}$ violates the rule above. It is customary to ignore the distinction between the meta-language and its semantics by referring to a particular consequence relation as \vDash and write $p \vDash q$ instead of $\langle p, q \rangle \in \vDash$.

A useful analogy with clausal logic is revealed by noting that the object language L establishes a Herbrand universe built up from the proposition symbols in L as constants and the connectives from L as function symbols. Consequence relations then correspond to Herbrand interpretations⁶ of the meta-language, whose rules can be easily transformed to clausal notation. The following terminology is borrowed from logic programming:

- (i) if all of P_1, \dots, P_n and Q are positive literals the rule is *definite*;
- (ii) if at least one of P_1, \dots, P_n is a negative literal and Q is a positive literal the rule is *indefinite*;

⁵This is not to say that the agent actually draws the conclusion when observing the premisses, but merely that she considers it one of the possible conclusions.

⁶These Herbrand interpretations are restricted to the meta-predicate \vDash ; \vDash is treated as a built-in predicate.

- (iii) if all of P_1, \dots, P_n are positive literals and Q is a negative literal the rule is a *denial*.⁷

For instance, all the rule systems of [Kraus *et al.*, 1990] are made up of definite rules only. Logic programming theory teaches us that the set of consequence relations satisfying a set of definite rules is closed under intersection. In contrast, the rule system \mathbf{R} characterised in [Lehmann and Magidor, 1992] contains one indefinite rule, *viz.* Rational Monotonicity. Consequently, the set of rational consequence relations is not closed under intersection.

3.2 Some properties of inductive consequence relations

After having stated the main definitions concerning consequence relations I will now list some properties generally obeyed by inductive consequence relations. In this section we will not distinguish between explanatory or confirmatory induction, and simply interpret $\alpha \prec \beta$ as ‘ β is a possible hypothesis given evidence α ’.

The first two rules state that the logical form of evidence and hypotheses is immaterial:

$$\text{Left Logical Equivalence} \quad \frac{\vDash \alpha \leftrightarrow \beta \quad , \quad \alpha \prec \gamma}{\beta \prec \gamma}$$

$$\text{Right Logical Equivalence} \quad \frac{\vDash \beta \leftrightarrow \gamma \quad , \quad \alpha \prec \beta}{\alpha \prec \gamma}$$

Left Logical Equivalence states, for instance, that if the evidence is expressed as a conjunction of ground facts, the order in which they occur is immaterial. From the viewpoint of practical induction algorithms this may embody a considerable simplification — however, the framework represented in this chapter is intended to provide a model for inductive reasoning in general rather than particular algorithms.⁸

The following two rules express principles well-known from philosophy of science:

$$\text{Verification} \quad \frac{\vDash \alpha \wedge \beta \rightarrow \gamma \quad , \quad \alpha \prec \beta}{\alpha \wedge \gamma \prec \beta}$$

$$\text{Falsification} \quad \frac{\vDash \alpha \wedge \beta \rightarrow \gamma \quad , \quad \alpha \prec \beta}{\alpha \wedge \neg \gamma \not\prec \beta}$$

⁷This exhausts all the possibilities: the case that at least one of P_1, \dots, P_n is a negative literal and Q is a negative literal can be rewritten to (i) or (ii).

⁸Practical algorithms establish a *function* from evidence to hypothesis rather than a relation, i.e. also Right Logical Equivalence would be invalidated by an induction algorithm (of all the logically equivalent hypotheses only one would be output).

In these two rules γ is a *prediction* made on the basis of hypothesis β and evidence α . Verification expresses that if such a prediction is indeed observed, hypothesis β remains a possible hypothesis, while if its negation is observed, β may be considered refuted according to Falsification.⁹ One might remark that typically the inductive hypothesis will entail the evidence, so that the first condition in the antecedent of Verification and Falsification may be simplified to $\models \beta \rightarrow \gamma$. However, this is only the case for certain approaches to explanatory induction; generally speaking inductive hypotheses, in particular those that are confirmed without being an explanation, may not contain all the information conveyed by the evidence. The formulation above represents the general case.

Falsification can be simplified in another sense, as shown by the following lemma.

LEMMA 1. *In the presence of Left Logical Equivalence, Falsification is equivalent with the following rule:*

$$\text{Consistency} \quad \frac{\alpha \not\vdash \beta}{\not\vdash \beta \rightarrow \neg\alpha}$$

Proof. To derive Falsification, suppose $\models \alpha \wedge \beta \rightarrow \gamma$, i.e. $\models \beta \rightarrow \neg(\alpha \wedge \neg\gamma)$, then by Consistency $\alpha \wedge \neg\gamma \not\vdash \beta$. To derive Consistency from Falsification, suppose $\alpha \not\vdash \beta$ and $\models \beta \rightarrow \neg\alpha$, i.e. $\models \alpha \wedge \beta \rightarrow \mathbf{false}$, then by Falsification $\alpha \wedge \mathbf{false} \not\vdash \beta$, and by Left Logical Equivalence $\alpha \not\vdash \beta$, a contradiction. ■

Falsification and Consistency rule out inconsistent evidence and hypotheses. The way inconsistent evidence is handled is merely a technicality, and we might have decided to treat it differently — for instance, Hempel’s entailment condition (H1) implies that in his framework inconsistent evidence confirms arbitrary hypotheses. The case of inconsistent hypotheses is different however: it is awkward to say, for instance, that arbitrary evidence induces an inconsistent hypothesis. Furthermore, in inductive concept learning often negative examples are included, that are not to be classified as belonging to the concept, which requires consistency of the induced rule. Also, the adoption of Consistency is the only way to treat explanatory and confirmatory induction in a unified way as regards the consistency of evidence and hypothesis.

In the presence of Consistency a number of other principles have to be formulated carefully. For instance, we have reflexivity only for consistent formulae.

⁹Notice that, contrary to the previous two rules, Verification and Falsification happen to be meaningful also when modelling the behaviour of an induction algorithm: Verification expresses that the current hypothesis should not be abandoned when the next observation is a predicted one (in the terminology of [Angluin and Smith, 1983] the algorithm is conservative), while Falsification expresses that the current hypothesis must be abandoned when the next observations run counter to the predictions of the algorithm (called consistency by [Angluin and Smith, 1983]). However, in the context of the present paper these are not the intended interpretations of the two rules.

In the light of Consistency a formula is consistent if it occurs in an inductive argument, either as evidence or as hypothesis, so we have the following weaker versions of reflexivity:¹⁰

$$\text{Left Reflexivity} \quad \frac{\alpha \llcorner \beta}{\alpha \llcorner \alpha}$$

$$\text{Right Reflexivity} \quad \frac{\alpha \llcorner \beta}{\beta \llcorner \beta}$$

If a consequence relation contains an argument $\alpha \llcorner \alpha$, this signals that α is consistent with the reasoner's background theory. We will call such an α *admissible* (with respect to the consequence relation), and use conditions of this form whenever we require consistency of evidence or hypothesis in a rule.¹¹

The final rule mentioned in this section is a variant of Verification that allows to add any prediction to the hypothesis rather than the evidence:

$$\text{Right Extension} \quad \frac{\vDash \alpha \wedge \beta \rightarrow \gamma \quad , \quad \alpha \llcorner \beta}{\alpha \llcorner \beta \wedge \gamma}$$

Further rules considered in this chapter are specific to either explanatory or confirmatory induction, and are therefore to be discussed in later sections.

3.3 The pragmatics of consequence relations

Before proceeding with the technical results of the chapter, we may spend a few thoughts on the exact nature of consequence relations. As defined above, a consequence relation is an extensional specification of the behaviour of a reasoning agent. The symbol \llcorner is introduced in order to reason about consequence relations and reasoning behaviour, and functions as a binary predicate in the metalanguage. Rules describing properties of \llcorner express boundaries of rationality of inductive reasoning: a consequence relation violating such a rule would be considered irrational.

Now suppose \mathbf{X} is a set of such rationality postulates, and let A be a set of inductive arguments. Clearly, by means of \mathbf{X} we could derive additional inductive arguments A' , the significance of which is: if an inductive agent accepts arguments A , and it behaves rationally according to \mathbf{X} , it should also accept arguments A' . For instance, if \mathbf{X} includes the rule of Verification, A contains the argument

`chad_is_black` \llcorner `crows_are_black`

and the background knowledge includes

¹⁰One might argue that induction is inherently non-reflexive if the hypothesis is to generalise the evidence. This point will be taken up in Section 6.

¹¹Readers with a background in inductive learning may interpret $\alpha \llcorner \alpha$ as 'hypothesis α does not cover any negative example'.

$$\models \text{crows_are_black} \rightarrow \text{chevy_is_black}$$

then A' contains the additional inductive argument

$$\text{chad_is_black} \wedge \text{chevy_is_black} \not\vdash \text{crows_are_black}$$

implying that if the agent would not accept the latter, it would behave irrationally (w.r.t. \mathbf{X}).

One may now ask: what is the smallest set of arguments containing A and satisfying the rules of \mathbf{X} ? Such a set, if it exists, would represent the closure of A under \mathbf{X} , denoted $A^{\mathbf{X}}$. The significance of such a closure is that if two agents start from the same set of arguments A , they cannot possibly disagree about any other argument if they both act rationally according to \mathbf{X} . Clearly, the existence of such a closure operation depends on the rules in \mathbf{X} . As has been remarked before, if all the rules in \mathbf{X} are definite (having only positive literals in their antecedents and consequent) the closure of A under \mathbf{X} is unique for arbitrary A . All the rule systems in [Kraus *et al.*, 1990] consist solely of definite rules.

However, the situation changes drastically if not all rules are definite. In particular, rule systems containing indefinite rules (having at least one negative literal in their antecedent and a positive literal as consequent) will not have an associated closure operation. An indefinite rule represents a rationality principle of the following kind: ‘if you accept this argument, you should accept at least one of those’. One example is Rational Monotonicity as studied in [Lehmann and Magidor, 1992]; other examples will be found in this chapter. The upshot of such rules is that even if two agents agree on the initial set of arguments they accept, they can disagree about some other arguments without violating the rationality postulates. Put differently, if we have two consequence relations both satisfying the rules of \mathbf{X} , and we take their intersection to find out on what arguments they agree, this intersection itself may not satisfy the rules of \mathbf{X} .

Lehmann and Magidor are not content with this indefiniteness: they define an operation of *rational closure* which selects, among the many supersets of A that satisfy \mathbf{X} , one that has certain desirable properties [Lehmann and Magidor, 1992, p. 33]. This seems to be motivated by the tacit assumption that *rationality postulates \mathbf{X} should lead to a closure operation $A \rightarrow A^{\mathbf{X}}$* . Such a closure operation can be seen as a consequence operation in the sense of Tarski [1936], mapping any set of premisses to the set of its consequences under some inference system. However, the notion of inference represented by such closure operations should be clearly distinguished from the notion of inference represented by consequence relations — if the latter operate on a meta-level, closure operations establish a *meta-meta-level*.

In this chapter we will not be concerned with inference on this meta-meta-level. This choice is motivated on methodological rather than technical grounds: I don’t believe that rationality postulates for induction necessarily establish an unequivocal closure operation. If we intersect two inductive consequence relations to see what two inductive agents agree upon, the resulting set of arguments may, as a

consequence relation, not satisfy some rationality postulate, but this is, it seems to me, just to be expected.

A further criticism of the meta-meta-level view is that it obscures the status of the consequence relation symbol. The relevant question on the meta-meta-level is: what expressions of the form $\alpha \sim \beta$ are entailed by a set of expressions of the same form? In other words: on the meta-meta-level \sim acts as a *connective*.¹² This is also apparent from Kraus, Lehmann and Magidor’s terminology: they call $\alpha \sim \beta$ a *conditional assertion*, a set of such conditional assertions is called a *conditional knowledge base*, and they ask themselves the question: *what does a conditional knowledge base entail?*, i.e. what other conditional assertions can be deduced from it? It seems to me that the meta-meta-level perspective is at odds with the meta-level perspective — indeed, it is my conjecture that the theory of entailment of conditional assertions can be developed without reference to an intermediate level of consequence relations.

I will now proceed with a technical analysis of the process of forming an explanatory hypothesis from evidence *à la* Peirce (Section 4) and the process of forming a confirmed hypothesis *à la* Hempel (Section 5).

4 EXPLANATORY INDUCTION

In this section we will study abstract properties and semantics for explanatory consequence relations. Throughout the section $\alpha \ll \beta$ is to be read as ‘evidence α is explained by hypothesis β ’ or ‘hypothesis β is a possible explanation of evidence α ’. What counts as a possible explanation will initially be left unspecified — the framework of consequence relations allows us to formulate abstract properties of hypothesis generation, without fixing a particular material definition. We will then single out a particular set of properties (the system **EM**) and characterise it semantically by means of strong explanatory structures.

4.1 Properties of explanatory induction

A natural requirement for explanatory induction is that every consistent hypothesis that entails the evidence counts as a possible explanation. As explained above the condition that hypothesis β be consistent is expressed by $\beta \ll \beta$, which gives us the following rule:

$$\text{Admissible Converse Entailment} \quad \frac{\vDash \beta \rightarrow \alpha \quad , \quad \beta \ll \beta}{\alpha \ll \beta}$$

Another requirement for explanations has been discussed above as (H5’): possible explanations may be logically strengthened, as long as they remain consistent.

¹²This raises the question why conditional assertions cannot be nested, as in $(\alpha \sim \beta) \sim \gamma$. Note that the answer to this question is perfectly clear on the meta-level, since this expression makes as little sense as, say, $(\alpha \vDash \beta) \vDash \gamma$.

This is expressed as follows:

$$\text{Admissible Right Strengthening} \quad \frac{\alpha \llcorner \beta \quad , \quad \vDash \gamma \rightarrow \beta \quad , \quad \gamma \llcorner \gamma}{\alpha \llcorner \gamma}$$

We may note that Admissible Converse Entailment can be derived from Admissible Right Strengthening if we assume Consistency and the following rule:

$$\text{Explanatory Reflexivity} \quad \frac{\alpha \llcorner \alpha \quad , \quad \neg \beta \not\llcorner \alpha}{\beta \llcorner \beta}$$

This rule represents a weakening of reflexivity especially tailored for explanatory induction. It is best understood by rewriting it into its contrapositive: from $\alpha \llcorner \alpha$ and $\beta \not\llcorner \beta$ infer $\neg \beta \llcorner \alpha$, which states that if β is inadmissible, i.e. too strong a statement with regard to the background knowledge, its negation $\neg \beta$ is so weak that it is explained by arbitrary admissible hypotheses α .

LEMMA 2. *In the presence of Consistency and Explanatory Reflexivity, Admissible Right Strengthening implies Admissible Converse Entailment.*

Proof. Suppose $\vDash \beta \rightarrow \alpha$, then by Consistency $\neg \alpha \not\llcorner \beta$. Suppose furthermore $\beta \llcorner \beta$, then by Explanatory Reflexivity $\alpha \llcorner \alpha$. The desired result follows by Admissible Right Strengthening. ■

While the rules above express properties of possible explanations, the following two rules concentrate on the evidence. The underlying idea is a basic principle in inductive learning: if the evidence is a set of instances of the concept, we can partition the evidence arbitrarily and find a single hypothesis that is an explanation of each subset of instances. This principle is established by the following two rules:¹³

$$\text{Incrementality} \quad \frac{\alpha \llcorner \gamma \quad , \quad \beta \llcorner \gamma}{\alpha \wedge \beta \llcorner \gamma}$$

$$\text{Convergence} \quad \frac{\vDash \alpha \rightarrow \beta \quad , \quad \alpha \llcorner \gamma}{\beta \llcorner \gamma}$$

LEMMA 3. *If \llcorner is a consequence relation satisfying Incrementality and Convergence, then $\alpha \wedge \beta \llcorner \gamma$ iff $\alpha \llcorner \gamma$ and $\beta \llcorner \gamma$.*

Proof. The *if* part is Incrementality, and the *only-if* part follows from Convergence. ■

¹³In previous work [Flach, 1995] Incrementality was called Additivity, and Convergence was called Incrementality. The terminology employed here better reflects the meaning of the rules.

Incrementality and Convergence are of considerable importance for computational induction, since they allow for an incremental approach. Incrementality states that pieces of evidence can be dealt with in isolation. Another way to say the same thing is that the set of evidence explained by a given hypothesis is conjunctively closed. By the rule of Consistency this set is consistent, which yields the following principle:

$$\text{Left Consistency} \quad \frac{\alpha \ll \beta}{\neg \alpha \not\ll \beta}$$

LEMMA 4. *In the presence of Right Reflexivity and Admissible Converse Entailment, Left Consistency implies Consistency.*

Proof. Suppose $\vDash \beta \rightarrow \neg \alpha$. Now, either $\beta \ll \beta$ or $\beta \not\ll \beta$; in the former case, $\neg \alpha \ll \beta$ by Admissible Converse Entailment, and we conclude by Left Consistency. In the latter case, we have $\delta \not\ll \beta$ for any δ by Right Reflexivity. ■

It follows that Left Consistency and Consistency are equivalent in the presence of Right Reflexivity, Admissible Converse Entailment, and Incrementality.

Convergence expresses a *monotonicity* property of induction, which can again best be understood by considering its contrapositive: a hypothesis that is rejected on the basis of evidence β cannot become feasible again when stronger evidence α is available. In other words: the process of rejecting a hypothesis is not defeasible (i.e. based on assumptions), but based on the evidence only. This is the analogue of the monotonicity property of deduction, which states that the process of *accepting* a consequence is non-defeasible (note that this property of deductive monotonicity can be obtained by reversing the implication in the first condition of Convergence).

LEMMA 5. *The combination of Verification and Convergence is equivalent with the following rule:*

$$\text{Predictive Convergence} \quad \frac{\vDash \alpha \wedge \gamma \rightarrow \beta \quad , \quad \alpha \ll \gamma}{\beta \ll \gamma}$$

Proof. To derive Predictive Convergence, suppose $\vDash \alpha \wedge \gamma \rightarrow \beta$ and $\alpha \ll \gamma$, then by Verification $\alpha \wedge \beta \ll \gamma$, and by Convergence $\beta \ll \gamma$.

Predictive Convergence implies Convergence, since $\vDash \alpha \rightarrow \beta$ implies $\vDash \alpha \wedge \gamma \rightarrow \beta$.

Predictive Convergence implies Verification, since $\vDash \alpha \wedge \beta \rightarrow \gamma$ implies $\vDash \alpha \wedge \beta \rightarrow \alpha \wedge \gamma$. ■

Predictive Convergence can be seen as a strengthening of Convergence, in the sense that β is not merely a weakening of evidence α , but can be any set of predictions. Note that Right Reflexivity is an instance of Predictive Convergence (put $\gamma = \beta$).

The final postulate we consider in this section expresses a principle well-known from algorithmic concept learning: if α represents the classification of an instance and β its description, then we may either induce a concept definition from examples of the form $\beta \rightarrow \alpha$, or we may add β to the background theory and induce from α alone. Since in our framework background knowledge is included implicitly, β is added to the hypothesis instead.

$$\text{Conditionalisation} \quad \frac{\alpha \ll \beta \wedge \gamma}{\beta \rightarrow \alpha \ll \gamma}$$

After having discussed various abstract properties of generation of explanatory hypotheses we now turn to the question of characterising explanatory induction semantically.

4.2 Strong explanatory consequence relations

As we have seen, Peirce's original idea was to define explanatory hypothesis generation as reversed deduction. I will amend Peirce's proposal in two ways. First, as explained above it is required that the hypothesis be consistent with respect to the background knowledge. Secondly, I reformulate reversed deduction as inclusion of deductive consequences. The main reason for the latter is that in this way the explanatory consequence relation is defined in terms of a property that is *preserved* by arguments (*viz.* explanatory power).

DEFINITION 6. An *explanation mechanism* is some consequence relation $\vdash\sim$. The *explanatory consequence relation* \ll defined by $\vdash\sim$ is defined as $\alpha \ll \beta$ iff $C_\sim(\alpha) \subseteq C_\sim(\beta) \subset L$, where $C_\sim(\alpha) = \{\gamma \mid \alpha \vdash\sim \gamma\}$. A *strong* explanatory consequence relation is defined by a monotonic explanation mechanism.

Thus, an explanation is required to have at least the same consequences under the explanation mechanism as the premiss it is obtained from, without becoming inconsistent. It should be noted that, in the general case, the conditions $C_\sim(\alpha) \subseteq C_\sim(\beta)$ and $\beta \vdash\sim \alpha$ are not equivalent.¹⁴ However, for monotonic explanation mechanisms they are, which provides us with the following 'Peircean' definition of strong explanatory consequence relations.

DEFINITION 7. A *strong explanatory structure* is a set $W \subseteq M$. The consequence relation it defines is denoted by \ll_W and is defined by: $\alpha \ll_W \beta$ iff (i) there is a $m_0 \in W$ such that $m_0 \vDash \beta$, and (ii) for every $m \in W$, $m \vDash \beta \rightarrow \alpha$.

The following system of rules will be proved to axiomatise strong explanatory structures.

DEFINITION 8. The system **EM** consists of the following rules: Admissible Right Strengthening, Explanatory Reflexivity, Incrementality, Predictive Convergence, Left Consistency, and Conditionalisation.

¹⁴ $C_\sim(\alpha) \subseteq C_\sim(\beta)$ implies $\beta \vdash\sim \alpha$ if $\vdash\sim$ is reflexive; $\beta \vdash\sim \alpha$ implies $C_\sim(\alpha) \subseteq C_\sim(\beta)$ if $\vdash\sim$ is transitive.

We note the following derived rules of **EM**: Convergence, Admissible Converse Entailment and Right Reflexivity (instances of Predictive Convergence) and Consistency (Lemma 4). The following derived rule will also prove useful.

LEMMA 9. *The following rule is a derived rule of **EM**:*

$$\text{Consistent Right Strengthening} \quad \frac{\alpha \llcorner \gamma \quad , \quad \neg \beta \not\llcorner \gamma}{\alpha \llcorner \beta \wedge \gamma}$$

Proof. Suppose $\neg \beta \not\llcorner \gamma$; since $\models \neg(\beta \wedge \gamma) \wedge \gamma \rightarrow \neg \beta$, we have $\neg(\beta \wedge \gamma) \not\llcorner \gamma$ by Predictive Convergence. Furthermore, suppose $\alpha \llcorner \gamma$, then by Right Reflexivity $\gamma \llcorner \gamma$, so by Explanatory Reflexivity we have $\beta \wedge \gamma \llcorner \beta \wedge \gamma$. We conclude by Admissible Right Strengthening. ■

Soundness of **EM** is easily checked.

LEMMA 10 (Soundness of **EM**). *Any strong explanatory consequence relation satisfies the rules of **EM**.*

Proof. Let $W \subseteq U$ be a strong explanatory structure; we need to demonstrate that \llcorner_W , as defined in Definition 7, satisfies the rules of **EM**.

Admissible Right Strengthening: if $\gamma \llcorner_W \gamma$, then some model in W satisfies γ . Furthermore, if $m \models \beta \rightarrow \alpha$ and $\models \gamma \rightarrow \beta$, then $m \models \gamma \rightarrow \alpha$.

Explanatory Reflexivity: we have that some model in W satisfies α , while not all models in W satisfy $\alpha \rightarrow \neg \beta$, i.e. there is a model in W satisfying $\alpha \wedge \beta$ and hence β .

Incrementality: if $m \models \gamma \rightarrow \alpha$ and $m \models \gamma \rightarrow \beta$ then $m \models \gamma \rightarrow (\alpha \wedge \beta)$.

Predictive Convergence: if $\models \alpha \wedge \gamma \rightarrow \beta$ and $m \models \gamma \rightarrow \alpha$ then $m \models \gamma \rightarrow \beta$.

Left Consistency: if some model in W satisfies β while all models in W satisfy $\beta \rightarrow \alpha$, then there is a model in W not satisfying $\beta \rightarrow \neg \alpha$.

Conditionalisation: trivial. ■

In order to prove completeness we build a strong explanatory structure W from a given consequence relation \llcorner satisfying the rules of **EM**, such that $\alpha \llcorner \beta$ iff $\alpha \llcorner_W \beta$. For non-empty explanatory relations the following construction is used:

$$W = \{m \in U \mid \text{for all } \alpha, \beta \text{ such that } \alpha \llcorner \beta : m \models \beta \rightarrow \alpha\}.$$

An empty explanatory relation signals inconsistent background knowledge, and is hence defined by the empty explanatory structure.

We need a few intermediate results. The following lemma states that every strong explanatory hypothesis is satisfiable in W .

LEMMA 11. *Let \llcorner be a consequence relation satisfying the rules of **EM**, and let W be defined as above. If $\alpha \llcorner \beta$ then there is a model $m \in W$ such that $m \models \beta$.*

Proof. Let $\alpha \vDash \beta$; we will prove that $\{\beta\} \cup \{\delta \rightarrow \gamma \mid \gamma \vDash \delta\}$ is satisfiable. Suppose not, then by compactness there is a finite $\Delta \subseteq \{\delta \rightarrow \gamma \mid \gamma \vDash \delta\}$ such that $\vDash \beta \rightarrow \neg\Delta$. Furthermore, since $\varphi \vDash \psi$ for any $\psi \rightarrow \varphi \in \Delta$, we have $\psi \rightarrow \varphi \vDash \mathbf{true}$ for any $\psi \rightarrow \varphi \in \Delta$ by Conditionalisation, $\Delta \vDash \mathbf{true}$ by Incrementality, and $\Delta \vDash \beta$ by Right Reflexivity and Admissible Right Strengthening. But then by Consistency $\not\vDash \beta \rightarrow \neg\Delta$, a contradiction. ■

Furthermore, we have that every inadmissible formula is unsatisfiable in W .

LEMMA 12. *Let \vDash be a non-empty consequence relation satisfying the rules of **EM**, and let W be defined as above. If $\gamma \not\vDash \gamma$ then γ is unsatisfiable in W .*

Proof. Let $\alpha \vDash \beta$, then $\mathbf{true} \vDash \mathbf{true}$ by Convergence and Left Reflexivity. Furthermore, if $\gamma \not\vDash \gamma$ then $\neg\gamma \vDash \mathbf{true}$ by Explanatory Reflexivity, hence $m \vDash \mathbf{true} \rightarrow \neg\gamma$ for every $m \in W$. ■

I will now show that W defines a consequence relation that is included in \vDash .

LEMMA 13. *Let \vDash be a non-empty consequence relation satisfying the rules of **EM**, and let W be defined as above. If $\alpha \vDash_W \beta$ then $\alpha \vDash \beta$.*

Proof. Suppose that $\alpha \not\vDash \beta$, we will show that either no model in W satisfies β , or there exists a model $m_0 \in W$ that does not satisfy $\beta \rightarrow \alpha$.

First of all, if $\beta \not\vDash \beta$ then β is unsatisfiable in W according to Lemma 12. In the remainder of the proof we will assume that $\beta \vDash \beta$. Define $\Gamma_0 = \{\neg\alpha\} \cup \{\delta \mid \delta \vDash \beta\}$; we will first show that Γ_0 is satisfiable. Suppose not, then by compactness there is a finite $\Delta \subseteq \{\delta \mid \delta \vDash \beta\}$ such that $\vDash \Delta \rightarrow \alpha$, i.e. $\vDash \beta \rightarrow (\Delta \rightarrow \alpha)$; by Admissible Converse Entailment $\Delta \rightarrow \alpha \vDash \beta$ (recall that $\beta \vDash \beta$). But by Incrementality $\Delta \vDash \beta$; using Incrementality and Convergence, we obtain $\alpha \vDash \beta$. Contradiction, so Γ_0 is satisfiable.

Let $m_0 \vDash \Gamma_0$; clearly $m_0 \not\vDash \alpha$ and, since $\beta \in \Gamma_0$, $m_0 \vDash \beta$. It remains to prove that m_0 is in W ; i.e., that for all φ, ψ such that $\varphi \vDash \psi$ we have $m_0 \vDash \psi \rightarrow \varphi$. Let $\varphi \vDash \psi$; if $\neg\beta \not\vDash \psi$, then by Consistent Right Strengthening $\varphi \vDash \psi \wedge \beta$, and by Conditionalisation $\psi \rightarrow \varphi \vDash \beta$; thus $\psi \rightarrow \varphi \in \Gamma_0$ and therefore $m_0 \vDash \psi \rightarrow \varphi$. On the other hand, if $\neg\beta \vDash \psi$ then by Conditionalisation and Convergence $\beta \rightarrow \neg\psi \vDash \mathbf{true}$, by Admissible Right Strengthening $\beta \rightarrow \neg\psi \vDash \beta$, and by Incrementality and Convergence $\neg\psi \vDash \beta$; thus $\neg\psi \in \Gamma_0$ and therefore $m_0 \not\vDash \psi$, hence $m_0 \vDash \psi \rightarrow \varphi$. ■

Armed with the previous three lemmas we can prove the completeness of **EM**.

THEOREM 14 (Representation theorem for strong explanatory consequence relations). *A consequence relation is strong explanatory iff it satisfies the rules of **EM**.*

Proof. The only-if part is Lemma 10. For the if part, let \vDash be an arbitrary non-empty consequence relation satisfying the rules of **EM**, and let

$$W = \{m \in U \mid \text{for all } \alpha, \beta \text{ such that } \alpha \vDash \beta : m \vDash \beta \rightarrow \alpha\}.$$

Suppose $\alpha \prec \beta$, then by the construction of W , $m \models \beta \rightarrow \alpha$ for all $m \in W$. Furthermore, by Lemma 11 there is a model in W satisfying β . We may conclude that $\alpha \prec_W \beta$. Conversely, if $\alpha \prec_W \beta$ then Lemma 13 proves that $\alpha \prec \beta$. We conclude that W defines a consequence relation that is exactly \prec .

For an empty consequence relation put $w = \emptyset$. ■

In this section we have studied axioms and semantic characterisations for explanatory induction. I have proposed a novel definition of explanatory hypothesis generation in terms of preservation of explanatory power with respect to an explanation mechanism. A representation theorem has been obtained for the special case of a monotonic explanation mechanism. Characterisation of explanatory induction with respect to other (e.g. preferential) explanation mechanisms is left as an open problem.

5 CONFIRMATORY INDUCTION

We will now switch from the explanatory (classification-oriented) viewpoint to the confirmatory (non-classificatory) perspective. Throughout this section $\alpha \prec \beta$ is to be read as ‘evidence α confirms hypothesis β ’. Our goals will be to find reasonable properties of \prec under this interpretation (for which we have a good starting point in Hempel’s adequacy conditions), and to characterise particular sets of properties by a suitable semantics.

5.1 *Properties of confirmatory induction*

In this section I will translate Hempel’s set of adequacy conditions for confirmation (Section 2.2) into rules for confirmatory consequence relations. The conditions will be slightly modified, in order to keep the treatment of inconsistent evidence and hypothesis in line with the explanatory case: inconsistent evidence does not confirm any hypothesis, and inconsistent hypotheses are not confirmed by any evidence.

Entailment condition (H1) is translated into two rules:

$$\begin{array}{l} \textbf{Admissible Entailment} \quad \frac{\models \alpha \rightarrow \beta \quad , \quad \alpha \prec \alpha}{\alpha \prec \beta} \\ \\ \textbf{Confirmatory Reflexivity} \quad \frac{\alpha \prec \alpha \quad , \quad \alpha \not\prec \neg \beta}{\beta \prec \beta} \end{array}$$

Admissible Entailment expresses that admissible evidence (i.e. evidence that is consistent with the background knowledge) confirms any of its consequences. In other words, consistent entailment is a special case of confirmation. Confirmatory Reflexivity is the confirmatory counterpart of Explanatory Reflexivity encountered

in the previous section. It is added as a separate rule since, in its original formulation, (H1) includes reflexivity as a special case. As with its explanatory counterpart, Confirmatory Reflexivity is best understood when considering its contrapositive: if β is inadmissible, i.e. too strong a statement with regard to the background knowledge, its negation $\neg\beta$ is so weak that it is confirmed by arbitrary admissible formulae α .

Consequence condition (H2) cannot be translated directly, since in the language of consequence relations as defined here we have no means to refer to a set of confirmed sentences. However, a translation of the special consequence condition (H2.1) and the conjunction condition (H2.3) will suffice:

$$\text{Right Weakening} \quad \frac{\vDash \beta \rightarrow \gamma \quad , \quad \alpha \ll \beta}{\alpha \ll \gamma}$$

$$\text{Right And} \quad \frac{\alpha \ll \beta \quad , \quad \alpha \ll \gamma}{\alpha \ll \beta \wedge \gamma}$$

LEMMA 15. *If \ll is a consequence relation satisfying Right And and Right Weakening, then $\alpha \ll \beta \wedge \gamma$ iff $\alpha \ll \beta$ and $\alpha \ll \gamma$.*

Proof. The if part is Right And, and the only-if part follows from Right Weakening. ■

Right Weakening expresses that any hypothesis entailed by a given hypothesis confirmed by α is also confirmed by α . Notice that Admissible Entailment is an instance of Right Weakening (put $\beta = \alpha$).

LEMMA 16. *The combination of Right Extension and Right Weakening is equivalent to the following rule:*

$$\text{Predictive Right Weakening} \quad \frac{\vDash \alpha \wedge \beta \rightarrow \gamma \quad , \quad \alpha \ll \beta}{\alpha \ll \gamma}$$

Proof. In order to derive Predictive Right Weakening, suppose $\vDash \alpha \wedge \beta \rightarrow \gamma$ and $\alpha \ll \beta$, then by Right Extension $\alpha \ll \beta \wedge \gamma$, and the result follows by Right Weakening.

Predictive Right Weakening implies Right Weakening, since $\vDash \beta \rightarrow \gamma$ implies $\vDash \alpha \wedge \beta \rightarrow \gamma$.

Predictive Right Weakening implies Right Extension, since $\vDash \alpha \wedge \beta \rightarrow \gamma$ implies $\vDash \alpha \wedge \beta \rightarrow \beta \wedge \gamma$. ■

In words, Predictive Right Weakening expresses that given a confirmatory argument, any predicted formula is confirmed by the same evidence. Notice that by putting $\gamma = \alpha$ in Predictive Right Weakening we obtain Left Reflexivity.

Right And states that the set of all confirmed hypotheses (interpreted as a conjunction) is itself confirmed. The combination of Right And and Right Weakening

implies Hempel's general consequence condition (H2): if E confirms every formula of a set K , then it also confirms the conjunction of the formulae in K (by Right And), and therefore also every consequence of this conjunction (by Right Weakening)¹⁵. It has already been remarked that Right And is probably too strong in the general case, if we have inconclusive evidence that is unable to choose between incompatible hypotheses. In this respect it is perhaps appropriate to point at a certain similarity between Right And and Right Extension: the latter rule requires γ to be predicted rather than being confirmed by α .

Like the general consequence condition (H2), general consistency condition (H3) cannot be translated directly into a rule, since we have no means to refer to the set of confirmed formulae. However, in the light of Right And the conjunction of the formulae in this set is itself confirmed, and therefore it is sufficient to formulate a rule expressing the special consistency condition (H3.1), which is the rule of Consistency previously encountered:

$$\text{Consistency} \quad \frac{\alpha \vDash \beta}{\not\vDash \beta \rightarrow \neg\alpha}$$

Condition (H3.2) expresses that for any formula β , if β is in the set of confirmed hypotheses then $\neg\beta$ is not. This principle is expressed by the following rule:

$$\text{Right Consistency} \quad \frac{\alpha \vDash \beta}{\alpha \not\vDash \neg\beta}$$

LEMMA 17. *In the presence of Admissible Entailment and Left Reflexivity, Right Consistency implies Consistency.*

Proof. Suppose $\vDash \beta \rightarrow \neg\alpha$, i.e. $\vDash \alpha \rightarrow \neg\beta$. Now, either we have $\alpha \vDash \alpha$, or else $\alpha \not\vDash \alpha$. In the former case, $\alpha \vDash \neg\beta$ by Admissible Entailment, and we conclude by Right Consistency. In the latter case, we have $\alpha \not\vDash \delta$ for any δ by Left Reflexivity. ■

Clearly, Consistency implies Right Consistency in the presence of Right And. As a corollary to Lemma 17, we have that Right Consistency and Consistency are equivalent in the presence of Left Reflexivity, Admissible Entailment, and Right And.

Finally, the equivalence condition for observations (H4) is translated into

$$\text{Left Logical Equivalence} \quad \frac{\vDash \alpha \leftrightarrow \beta \quad , \quad \alpha \vDash \gamma}{\beta \vDash \gamma}$$

We now turn to the question of devising a meaningful semantics for Hempel's conditions as re-expressed in our framework of inductive consequence relations.

¹⁵This holds only for finite K , an assumption that I will make throughout.

5.2 Simple confirmatory structures

It has been suggested in Section 2.4 that a semantics for confirmatory reasoning be expressed in terms of satisfaction by an appropriately constructed model or set of models. More precisely, a confirmatory semantics is conceived as one in which certain *regular* models are constructed from the premisses, such that a hypothesis is confirmed if it is true in all those regular models. Since this requires some completeness assumptions regarding the evidence we may call this a *closed* confirmatory semantics. In Section 5.4 I will consider a variant which relaxes the assumptions regarding the evidence (*open* confirmatory semantics).

DEFINITION 18. A *confirmatory structure* is a triple $W = \langle S, [\cdot], \llbracket \cdot \rrbracket \rangle$, where S is a set of *semantic objects*, and $[\cdot] : L \rightarrow 2^S$ and $\llbracket \cdot \rrbracket : L \rightarrow 2^S$ are functions mapping formulae to sets of semantic objects. The closed confirmatory consequence relation defined by W is given by: $\alpha \vDash_W \beta$ iff (i) $[\alpha] \neq \emptyset$, and (ii) $[\alpha] \subseteq \llbracket \beta \rrbracket$.

Intuitively, $[\alpha]$ denotes the set of regular models constructed from premisses α , each of which should satisfy hypothesis β . A similar semantics has been considered by Bell [1991], who calls it a pragmatic model. There are however two differences between Bell's approach and mine. First, in order to rule out inconsistent premisses I have added condition (i) in the definition of \vDash_W . Furthermore, I allow the possibility that some of the regular models may not satisfy the premisses (i.e. $[\alpha] \not\subseteq \llbracket \alpha \rrbracket$). The rationale of this possibility is further discussed in Section 6. However, the characterisation of such anti-reflexive logics of confirmation is left as an open problem, and the results obtained below are based on the assumption that $[\alpha] \subseteq \llbracket \alpha \rrbracket$ for all $\alpha \in L$. These results differ from those reported in [Kraus *et al.*, 1990; Bell, 1991] because of the additional condition (i), as will be detailed in the present section and the next.

Bell proves [1991, Theorem 4.8] that his pragmatic models, including the condition $[\alpha] \subseteq \llbracket \alpha \rrbracket$, are axiomatised by the rule system **B** consisting of Reflexivity, Right Weakening, and Right And, if one additionally assumes that $\llbracket \cdot \rrbracket$ is well-behaved with respect to the logical connectives and logical entailment:

$$\begin{aligned} \llbracket \alpha \wedge \beta \rrbracket &= \llbracket \alpha \rrbracket \cap \llbracket \beta \rrbracket \\ \llbracket \neg \alpha \rrbracket &= S - \llbracket \alpha \rrbracket \\ \llbracket \alpha \rrbracket &= S \text{ iff } \vDash \alpha \end{aligned}$$

Let us call confirmatory structures which satisfy these conditions, as well as $[\alpha] \subseteq \llbracket \alpha \rrbracket$ for all $\alpha \in L$, *simple* confirmatory structures, defining simple (closed) confirmatory consequence relations. I will now demonstrate that simple confirmatory structures are axiomatised by the following rule system.

DEFINITION 19. The system **CS** consists of the following rules: Predictive Right Weakening, Right And, and Right Consistency.

Derived rules of **CS** include Right Weakening and Right Extension (Lemma 16), Left Reflexivity (an instance of Predictive Right Weakening), Admissible Entailment (an instance of Right Weakening), and Consistency (Lemma 17).

The soundness of the rules of **CS** is easily proved.

LEMMA 20 (Soundness of CS). *Any simple closed confirmatory consequence relation satisfies the rules of CS.*

Proof. For Predictive Right Weakening, first of all we have $[\alpha] \subseteq \llbracket \alpha \rrbracket$ for all $\alpha \in L$, hence $[\alpha] \cap \llbracket \beta \rrbracket \subseteq \llbracket \alpha \rrbracket \cap \llbracket \beta \rrbracket$ for all $\alpha, \beta \in L$. Now suppose $\models \alpha \wedge \beta \rightarrow \gamma$, then $\llbracket \alpha \rrbracket \cap \llbracket \beta \rrbracket \subseteq \llbracket \gamma \rrbracket$, hence $[\alpha] \cap \llbracket \beta \rrbracket \subseteq \llbracket \gamma \rrbracket$. Furthermore suppose $\alpha \llcorner_W \beta$, then $\emptyset \subset [\alpha] \subseteq \llbracket \beta \rrbracket$, hence $[\alpha] \cap \llbracket \beta \rrbracket = [\alpha] \subseteq \llbracket \gamma \rrbracket$, and we conclude $\alpha \llcorner_W \gamma$.

For Right And, if $[\alpha] \subseteq \llbracket \beta \rrbracket$ and $[\alpha] \subseteq \llbracket \gamma \rrbracket$ then $[\alpha] \subseteq \llbracket \beta \rrbracket \cap \llbracket \gamma \rrbracket = \llbracket \beta \wedge \gamma \rrbracket$.

For Right Consistency, if $\emptyset \subset [\alpha] \subseteq \llbracket \beta \rrbracket$ then $[\alpha] \cap \llbracket \beta \rrbracket \neq \emptyset$, hence $[\alpha] \not\subseteq \llbracket \neg \beta \rrbracket$. ■

In order to prove completeness we need to build, for a given consequence relation, a simple confirmatory structure W that defines exactly that relation. The key concept is the following.

DEFINITION 21. Let \llcorner be a confirmatory consequence relation. The model $m \in U$ is said to be *normal for α* iff for all β in L such that $\alpha \llcorner \beta$, $m \models \beta$.

For admissible formulae, normal models will play the role of the regular models in the simple confirmatory structure we are building (remember that, given a consequence relation \llcorner , a formula $\alpha \in L$ is admissible iff $\alpha \llcorner \alpha$).

LEMMA 22. *Suppose a consequence relation \llcorner satisfies Right Weakening and Right And, and let α be an admissible formula. All normal models for α satisfy β iff $\alpha \llcorner \beta$.*

Proof. The if part follows from Definition 21.

For the only-if part, suppose $\alpha \llcorner \alpha$ and $\alpha \not\llcorner \beta$; I will show that there is a normal model for α that does not satisfy β . Let $\Gamma_0 = \{\neg \beta\} \cup \{\delta \mid \alpha \llcorner \delta\}$; it suffices to show that Γ_0 is satisfiable. Suppose not, then by compactness there is a finite $\Delta \subseteq \{\delta \mid \alpha \llcorner \delta\}$ such that $\models \Delta \rightarrow \beta$, i.e. $\models \alpha \rightarrow (\Delta \rightarrow \beta)$; by Right Weakening $\alpha \llcorner \Delta \rightarrow \beta$. But by Right And $\alpha \llcorner \Delta$; using Right And and Right Weakening we obtain $\alpha \llcorner \beta$, a contradiction. ■

In the standard treatment of inconsistent premisses they have all formulae as consequences, hence no normal models — since in our treatment inconsistent premisses confirm no hypothesis and thus have all models in U as normal models, we have to treat them as a separate case. Given a consequence relation \llcorner , let $W = \langle U, [\cdot], \llbracket \cdot \rrbracket \rangle$ be defined as follows:

1. U is the set of models of L under consideration;
2. $[\alpha] = \{m \in U \mid m \text{ is a normal model for } \alpha\}$ if α is admissible, and \emptyset otherwise;
3. $\llbracket \alpha \rrbracket = \{m \in U \mid m \models \alpha\}$.

The following completeness result demonstrates that the consequence relation defined by W coincides with the original one if the latter satisfies the rules of **CS**.

THEOREM 23 (Representation theorem for simple confirmatory consequence relations). *A consequence relation is simple confirmatory iff it satisfies the rules of **CS**.*

Proof. The only-if part is Lemma 20. For the if part, let \llcorner be a consequence relation satisfying the rules of **CS** and let W be defined as above. We will prove that $\alpha \llcorner \beta$ iff $\alpha \llcorner_W \beta$, i.e. \llcorner is simple confirmatory.

First suppose that $\alpha \llcorner \beta$, then by Left Reflexivity $\alpha \llcorner \alpha$ and by Right Consistency $\alpha \not\llcorner \neg\alpha$, so the proof of Lemma 22 constructs a model m normal for α , hence $[\alpha] \neq \emptyset$. Furthermore any normal model for α satisfies β by Definition 21, hence $[\alpha] \subseteq \llbracket \beta \rrbracket$. We conclude that $\alpha \llcorner_W \beta$. Now suppose that $\alpha \llcorner_W \beta$, then $[\alpha] \neq \emptyset$ hence α is admissible by the definition of W . Furthermore we have $[\alpha] \subseteq \llbracket \beta \rrbracket$, i.e. every model normal for α satisfies β , and the conclusion follows by Lemma 22. ■

One may note that two of the rules obtained in the previous section have not been mentioned above, *viz.* Confirmatory Reflexivity and Left Logical Equivalence. Each of these rules poses additional restrictions on simple confirmatory structures:

- | | |
|----------------------------|---|
| (Left Logical Equivalence) | If $\models \alpha \leftrightarrow \beta$, then $[\alpha] = [\beta]$. |
| (Confirmatory Reflexivity) | If $\llbracket \beta \rrbracket \neq \emptyset$, then $[\beta] \neq \emptyset$. |

Additional properties may be obtained by being more explicit about the construction of regular models. The ILP methods referred to in Section 2.4 suggest to take the truth-minimal Herbrand model(s) of the evidence as the regular model(s). In the analysis of nonmonotonic reasoning it is customary to abstract this into a preference ordering on the set of models, such that the regular models are the minimal ones under this ordering. We will work this out in the next section.

5.3 Preferential confirmatory consequence relations

The main result of this section concerns an adaptation of Kraus *et al.*'s preferential semantics, recast as a confirmatory semantics. In other words, the regular semantic objects are those that are minimal with respect to a fixed preference ordering.

DEFINITION 24. A *preferential structure* is a triple $W = \langle S, l, < \rangle$ where S is a set of *states*, $l : S \rightarrow U$ is a function that labels every state with a model, and $<$ is a strict partial order¹⁶ on S , called the *preference ordering*, that is smooth.¹⁷

¹⁶I.e. $<$ is irreflexive and transitive.

¹⁷I.e. for any $S' \subseteq S$ and for any $s \in S'$, or there is a $t \in S'$ such that $t < s$ and t is minimal in S' . This condition is satisfied if $<$ does not allow infinite descending chains.

W defines a *preferential confirmatory* structure $\langle S, [\cdot], \llbracket \cdot \rrbracket \rangle$ and a preferential confirmatory consequence relation as follows: $\llbracket \alpha \rrbracket = \{s \in S \mid l(s) \models \alpha\}$, and $[\alpha] = \{s \in \llbracket \alpha \rrbracket \mid \forall s' \in S : s' < s \rightarrow s' \notin \llbracket \alpha \rrbracket\}$.

Note that preferential confirmatory structures are simple confirmatory structures, and that they also satisfy the conditions associated above with Left Logical Equivalence and Confirmatory Reflexivity (by the smoothness condition, if $s \in \llbracket \alpha \rrbracket$ then either $s \in [\alpha]$ or there is a $t < s$ such that $t \in [\alpha]$ — hence $\llbracket \alpha \rrbracket \neq \emptyset$ implies $[\alpha] \neq \emptyset$).

In comparison with the preferential semantics of [Kraus *et al.*, 1990], the only difference is that in a preferential confirmatory argument the evidence is required to be satisfiable, in order to guarantee the validity of Consistency. The intermediate semantic level of *states* is mainly needed for technical reasons, and can be identified with the set of models the reasoning agent considers possible in that epistemic state.

The following set of rules will be proved to axiomatise preferential confirmatory consequence relations.

DEFINITION 25. The system **CP** consists of the rules of **CS**, Confirmatory Reflexivity, Left Logical Equivalence, plus the following rules:

$$\begin{array}{l} \textbf{Left Or} \qquad \frac{\alpha \vDash \gamma \quad , \quad \beta \vDash \gamma}{\alpha \vee \beta \vDash \gamma} \\ \\ \textbf{Strong Verification} \qquad \frac{\alpha \vDash \gamma \quad , \quad \alpha \vDash \beta}{\alpha \wedge \gamma \vDash \beta} \end{array}$$

The first rule can be seen as a variant of Convergence, which is clearly invalid in the general case: if we weaken the evidence, there will presumably come a point where the evidence no longer confirms the hypothesis. However, Left Or states that pieces of confirming evidence for a hypothesis can be weakened by taking their disjunction. The second rule is a variant of Verification, which states that a predicted formula γ can be added to confirming evidence α for hypothesis β . Strong Verification states that this is also allowed when γ is confirmed by α . The way in which Strong Verification strengthens Verification is very similar to the way Right And strengthens Right Extension. The underlying intuition is that the evidence is strong enough to have all confirmations ‘point in the same direction’, as it were.¹⁸

LEMMA 26 (Soundness of CP). *Any preferential confirmatory consequence relation satisfies the rules of CP.*

Proof. The proof only needs to be carried out for the two new rules.

¹⁸In the context of nonmonotonic reasoning, Strong Verification is known as Cautious Monotonicity. For the purposes of this paper I prefer to use the first name, which expresses more clearly the underlying intuition in the present context.

For Left Or, note that $\llbracket \alpha \vee \beta \rrbracket = \llbracket \alpha \rrbracket \cup \llbracket \beta \rrbracket$; thus, if $\llbracket \alpha \rrbracket$ and $\llbracket \beta \rrbracket$ are non-empty then so is $\llbracket \alpha \vee \beta \rrbracket$, hence $[\alpha \vee \beta] \neq \emptyset$. Furthermore, $[\alpha \vee \beta] \subseteq [\alpha] \cup [\beta] \subseteq \llbracket \alpha \vee \beta \rrbracket$, since a state cannot be minimal in $\llbracket \alpha \vee \beta \rrbracket$ without being minimal in at least one of $\llbracket \alpha \rrbracket$ and $\llbracket \beta \rrbracket$.

For Strong Verification, suppose that $[\alpha] \neq \emptyset$, $[\alpha] \subseteq \llbracket \gamma \rrbracket$ and $[\alpha] \subseteq \llbracket \beta \rrbracket$ — clearly, $\llbracket \alpha \wedge \gamma \rrbracket$ is non-empty, hence $[\alpha \wedge \gamma] \neq \emptyset$. Now, let $s \in [\alpha \wedge \gamma]$, then $s \in \llbracket \alpha \rrbracket$; I will prove that $s \in [\alpha]$. Suppose not, then there is a $t \in \llbracket \alpha \rrbracket$ such that $t < s$ and $t \in [\alpha] \subseteq \llbracket \gamma \rrbracket$, hence $t \in \llbracket \alpha \wedge \gamma \rrbracket$. But this contradicts the minimality of s in $\llbracket \alpha \wedge \gamma \rrbracket$, hence $s \in [\alpha]$ and thus $s \in \llbracket \beta \rrbracket$. ■

In order to prove completeness, we need to build a preferential structure W from a given consequence relation \llcorner satisfying the rules of **CP**, such that $\alpha \llcorner \beta$ iff $\llcorner_W \beta$. This structure $W = \langle S, l, < \rangle$ is constructed as follows:

1. $S = \{ \langle m, \alpha \rangle \mid \alpha \text{ is an admissible formula, and } m \text{ is a normal model for } \alpha \}$;
2. $l(\langle m, \alpha \rangle) = m$;
3. $\langle m, \alpha \rangle < \langle n, \beta \rangle$ iff $\alpha \vee \beta \llcorner \alpha$ and $m \not\models \beta$.

Thus, states are pairs of admissible formulae and normal models. The labelling function simply maps a state to the model it contains. Condition (3) defines the preference ordering between states: note that $\beta \llcorner \alpha$ is a special case of $\alpha \vee \beta \llcorner \alpha$ by means of Left Or, and the fact that α is admissible. The condition $m \not\models \beta$ is added to make the ordering irreflexive; note that as a consequence any $\langle m, \alpha \rangle \in S$ is minimal in $\llbracket \alpha \rrbracket$.

The main difference between the preferential consequence relations of Kraus *et al.* and my preferential confirmatory consequence relations is the way unsatisfiable formulae are treated. In Kraus *et al.*'s system **P** unsatisfiable formulae are characterised by the fact that they have every formula in L as a plausible consequence, which means that they don't have normal models. In my framework, unsatisfiable formulae confirm no hypotheses, and have all models in U as normal models. In both cases, the structure W that is used to prove completeness contains only satisfiable formulae in its states. This means that we can replicate most of Kraus *et al.*'s results about the structure W .

PROPOSITION 27.

1. [Kraus *et al.*, 1990, Lemma 5.13] *The relation $<$ is a strict partial order.*
2. [Kraus *et al.*, 1990, Lemma 5.15] *The relation $<$ is smooth: for any $s \in \llbracket \alpha \rrbracket$, either s is minimal in $\llbracket \alpha \rrbracket$ or there exists a state $t < s$ minimal in $\llbracket \alpha \rrbracket$.*
3. [Kraus *et al.*, 1990, Lemma 5.11] *If $\alpha \vee \beta \llcorner \alpha$ and m is a normal model for α that satisfies β , then m is a normal model for β .*
4. [Kraus *et al.*, 1990, Lemma 5.14] *$\langle m, \alpha \rangle \in \llbracket \beta \rrbracket$ iff $m \models \beta$ and $\alpha \vee \beta \llcorner \alpha$.*

The first two statements express that W is a preferential structure. The remaining two are used in the proof of the following lemma.

LEMMA 28. *Let \llcorner be a consequence relation satisfying the rules of **CP**, and let W be defined as above. If $\alpha \llcorner \beta$ then $\alpha \llcorner_W \beta$.*

Proof. First suppose that $\alpha \llcorner \beta$, then by Left Reflexivity $\alpha \llcorner \alpha$ and by Right Consistency $\alpha \not\llcorner \neg\alpha$, so the proof of Lemma 22 constructs a model m normal for α , hence $[\alpha] \neq \emptyset$. Furthermore suppose $s = \langle n, \gamma \rangle \in [\alpha]$, then γ is an admissible formula, n is a normal model for γ that satisfies α , and $\gamma \vee \alpha \llcorner \gamma$ by Proposition 27 (4). By Proposition 27 (3) n is a normal model for α and thus satisfies β by Definition 21, hence $s \in \llbracket \beta \rrbracket$. We conclude that $\emptyset \subset [\alpha] \subseteq \llbracket \beta \rrbracket$, hence $\alpha \llcorner_W \beta$. ■

The following lemma proves the converse of Lemma 28, and completes the proof of the representation theorem.

LEMMA 29. *Let \llcorner be a consequence relation satisfying the rules of **CP**, and let W be defined as above. If $\alpha \llcorner_W \beta$ then $\alpha \llcorner \beta$.*

Proof. Suppose $\alpha \llcorner_W \beta$, i.e. $\emptyset \subset [\alpha] \subseteq \llbracket \beta \rrbracket$; we will first prove that α is admissible. Let $\langle n, \gamma \rangle \in [\alpha]$, then $\gamma \llcorner \gamma$ and n is a normal model for γ . If $\alpha \not\llcorner \alpha$, then by Confirmatory Reflexivity $\gamma \llcorner \neg\alpha$ and thus $n \models \neg\alpha$, contradicting the assumption that $\langle n, \gamma \rangle \in [\alpha]$ — so α is admissible. Furthermore, given any model m normal for α , $\langle m, \alpha \rangle \in [\alpha] \subseteq \llbracket \beta \rrbracket$, hence m satisfies β , and the conclusion follows by Lemma 22. ■

We may now summarise.

THEOREM 30 (Representation theorem for preferential confirmatory consequence relations). *A consequence relation is preferential confirmatory iff it satisfies the rules of **CP**.*

Proof. The only-if part is Lemma 26. For the if part, let \llcorner be a consequence relation satisfying the rules of **CP** and let W be defined as above. Lemma 28 and Lemma 29 prove that $\alpha \llcorner \beta$ iff $\alpha \llcorner_W \beta$, i.e. \llcorner is preferential confirmatory. ■

We end this section by noting that **CP** represents a more restrictive form of reasoning than Kraus *et al.*'s system **P**, even though neither system entails the other (Reflexivity, a rule of **P**, is invalid in **CP**, while Consistency, a rule of **CP**, is invalid in **P**). This is so because from every preferential consequence relation \llcorner we can construct a preferential confirmatory consequence relation by removing all arguments $\{\alpha \llcorner \beta \mid \alpha, \beta \in L \text{ and } \alpha \llcorner \neg\alpha\}$, i.e. all arguments with a left-hand side that is inconsistent with respect to the background knowledge. For further details the reader is referred to [Flach, 1998].

5.4 Weak confirmatory consequence relations

Any semantics that is to obey rules like Right And and Strong Verification must be based on completeness assumptions with regard to the evidence. On the other hand, such strong assumptions cannot be made for all induction tasks. It seems reasonable, then, to investigate also an alternative approach, in which a confirmed hypothesis is required to be true in *some* of the regular models.¹⁹ It is not difficult to see that such an alternative semantics, based on some notion of consistency, would invalidate both Right And and Strong Verification. On the other hand, by not making completeness assumptions one may again have the desirable property of Convergence.

DEFINITION 31. Let $W = \langle S, [\cdot], \llbracket \cdot \rrbracket \rangle$ be a confirmatory structure. The open confirmatory consequence relation defined by W is given by: $\alpha \llcorner_W \beta$ iff $[\alpha] \cap \llbracket \beta \rrbracket \neq \emptyset$.

We will characterise an extreme form of open confirmatory relations, which arises when $[\cdot]$ is identified with $\llbracket \cdot \rrbracket$, which in turn is well-behaved with respect to the connectives and entailment.

DEFINITION 32. A *classical confirmatory structure* is a simple confirmatory structure $\langle S, \llbracket \cdot \rrbracket, \llbracket \cdot \rrbracket \rangle$. A consequence relation is called *weak confirmatory* iff it is the open consequence relation defined by a classical confirmatory structure.

From this definition it is clear that weak confirmatory consequence relations satisfy both Right Weakening and Left Weakening (i.e. Convergence), as well as Consistency. One additional rule is needed.

DEFINITION 33. The system **CW** consists of the following rules: Predictive Convergence, Predictive Right Weakening, Consistency, and

$$\text{Disjunctive Rationality} \quad \frac{\alpha \vee \beta \llcorner \gamma \quad , \quad \beta \not\llcorner \gamma}{\alpha \llcorner \gamma}$$

Disjunctive Rationality has not been considered in this chapter before. The name has been borrowed from Kraus *et al.*, who identify it as a valid principle of plausible reasoning. In the context of confirmatory reasoning, Disjunctive Rationality is a rather strong rule, which states that if a hypothesis is confirmed by disjunctive observations it is confirmed by at least one of the disjuncts. The following theorem proves the equivalence of weak confirmatory structures and the system **CW**.

THEOREM 34 (Representation theorem for weak confirmatory consequence relations). *A consequence relation \llcorner is weak confirmatory iff it satisfies the rules of **CW**.*

Proof. The only-if part involves demonstrating that \llcorner_W , as defined in Definition 32, satisfies the rules of **CW**, which is trivial.

¹⁹This is sometimes called *credulous* inference, in contrast with *skeptical* inference which requires truth in all regular models.

For the if part, let \llcorner be an arbitrary consequence relation satisfying the rules of **CW**, and consider the weak confirmatory structure $W = \langle S, \llbracket \cdot \rrbracket, \llbracket \cdot \rrbracket \rangle$ with:

$$S = \{m \in U \mid \text{for all } \alpha, \beta \in L \text{ such that } m \models \alpha \wedge \beta : \alpha \llcorner \beta\}$$

$$\llbracket \alpha \rrbracket = \{m \in S \mid m \models \alpha\}$$

We will prove that $\alpha \llcorner \beta$ iff $\alpha \llcorner_W \beta$. The if part follows directly from the construction of W . Suppose that $\alpha \llcorner \beta$, we will show that there exists a model $m_0 \in S$ that satisfies $\alpha \wedge \beta$.

Define $\Gamma_0 = \{\alpha\} \cup \{\delta \mid \neg\delta \not\llcorner \beta\}$; we will first show that Γ_0 is satisfiable. Suppose not, then by compactness there is a finite $\Delta \subseteq \{\delta \mid \neg\delta \not\llcorner \beta\}$ such that $\models \neg(\Delta \wedge \alpha)$, i.e. $\models \beta \rightarrow \neg(\Delta \wedge \alpha)$; by Consistency $\Delta \wedge \alpha \not\llcorner \beta$. Furthermore, since $\neg\delta \not\llcorner \beta$ for $\delta \in \Delta$, we have $\neg\Delta \not\llcorner \beta$ by Disjunctive Rationality and $\neg\Delta \wedge \alpha \not\llcorner \beta$ by Convergence. Combining this with $\Delta \wedge \alpha \not\llcorner \beta$ we obtain $(\neg\Delta \wedge \alpha) \vee (\Delta \wedge \alpha) \not\llcorner \beta$ by Disjunctive Rationality and $\alpha \not\llcorner \beta$ by Incrementality. Contradiction, so Γ_0 is satisfiable.

Let $m_0 \models \Gamma_0$; clearly $m_0 \models \alpha$ and, since by Consistency $\beta \in \Gamma_0$, $m_0 \models \beta$. It remains to prove that $m_0 \in S$; i.e., that for all φ, ψ such that $m_0 \models \varphi \wedge \psi$ we have $\varphi \llcorner \psi$. Let $m_0 \models \varphi \wedge \psi$, then $\neg(\varphi \wedge \psi) \notin \Gamma_0$, hence $\varphi \wedge \psi \llcorner \beta$; by Predictive Right Weakening $\varphi \wedge \psi \llcorner \psi \wedge \beta$, by Convergence $\varphi \llcorner \psi \wedge \beta$, and by Right Weakening $\varphi \llcorner \psi$. ■

The system **CW** thus provides an axiomatisation of the relation of logical compatibility.

In this section we have studied abstract properties and semantics for confirmatory induction. In the first part of this analysis I have demonstrated that Hempel's original conditions axiomatise a rather general form of confirmatory inference, characterised by truth in the regular models of the evidence. A close link with nonmonotonic or plausible reasoning is obtained by identifying the regular models with the minimal models under a preference ordering. Finally, I have proposed a more liberal form of confirmatory reasoning based on some notion of consistency, which is more appropriate if the evidence cannot be considered complete. This more liberal form of confirmatory reasoning invalidates the strong rule of Right And. A new representation theorem has been obtained for the extreme form of logical compatibility. Open problems include dropping the condition that regular models be models of the premisses, and finding more meaningful forms of open confirmatory reasoning.

6 DISCUSSION

In this chapter I have combined and extended old and recent work in philosophy, logic, and machine learning, in an attempt to gain more insight in induction as a reasoning process. I believe that the approach followed has implications for each of these three disciplines, which I will address separately below.

6.1 *Philosophy*

Induction is one of the traditional problems of philosophy. In spite of this it is perhaps also one of the least understood, and certainly one that hasn't been satisfyingly solved. From the perspective from which this chapter is written there is not one, but two 'problems of induction': one concerned with the generation of possible hypotheses, and one concerned with the justification of accepted hypotheses. Traditionally, philosophers have been concerned with the justification problem:

'Why is a single instance, in some cases, sufficient for a complete induction, while in others, myriads of concurring instances, without a single exception known or presumed, go such a very little way towards establishing an universal proposition? Whoever can answer this question knows more of the philosophy of logic than the wisest of the ancients, and has solved the problem of induction.'

[Mill, 1843, Book III, Chapter VIII, p. 314]

No attempt has been made, in the present chapter, to solve *this* problem of induction. As I have argued in Section 2.1, I don't think the justification problem manifests itself exclusively with inductive generalisations, but rather with all forms of nondeductive reasoning.

Furthermore, I don't think that the justification problem is a problem of logic. What I perceive as the *logical problem of induction*, and what has been the central problem of this chapter, is the problem of finding a sufficiently accurate description, in logical terms, of the process of inductive hypothesis generation. As Hanson observes, this logical problem of induction has been mostly ignored:

'Logicians of science have described how one might set out reasons in support of an hypothesis once it is proposed. They have said little about the conceptual considerations pertinent to the initial proposal of an hypothesis. There are two exceptions: Aristotle and Peirce. When they discussed what Peirce called "retroduction",²⁰ both recognized that the proposal of an hypothesis is often a reasonable affair. (...)

Neither Aristotle nor Peirce imagined himself to be setting out a manual to help scientists make discoveries. There could be no such manual. Nor were they discussing the psychology of discoverers, or the sociology of discovery. There are such discussions, but they are not logical discussions. Aristotle and Peirce were doing logic. They examined characteristics of the reasoning behind the original suggestion of certain hypotheses.'

[Hanson, 1958, pp. 1073–4]

The first philosophical contribution of this chapter, then, has been to reinforce the point that inductive hypothesis generation is a logical matter (and hypothesis

²⁰Peirce's translation of Aristotle's term $\alpha\pi\alpha\gamma\omega\gamma\eta$ — only later Peirce introduced the term 'abduction'.

selection is not). The second contribution lies in the distinction I have drawn between explanatory induction and confirmatory induction. More generally, I believe that a logical characterisation of inductive hypothesis generation is impossible unless one takes into account the *goal* which the hypothesis is intended to fulfil. Possible goals include providing classifications like in concept learning, or making implicit regularities explicit like in knowledge discovery in databases. Different goals lead to different forms of induction with different logical characteristics. For this reason I have taken a somewhat relativistic viewpoint, by setting up a general framework in which various logics of induction can be formulated, analysed, and compared, instead of fixing a particular logic of induction.

Two main families of logics of induction have been singled out, one portraying induction as explanation-preserving reasoning, the other as inference of confirmed hypotheses. As these families should be taken as starting points for further technical research, let us consider a number of possible improvements here. First of all, it has been remarked earlier that the restriction to a propositional object language L is counterintuitive in the context of induction, where the distinction between statements about individuals and statements about sets of individuals appears to be crucial. Upgrading the results of this chapter to predicate logic is certainly an important open problem. On the other hand, the propositional analysis of this chapter is not meaningless for the predicate logic case. First of all, in finite domains statements of predicate logic can be encoded in propositional logic.²¹ Furthermore, I would expect each of the predicate logic rule systems to include the corresponding propositional rules, and the predicate logic semantics to be refinements of the propositional semantics.

A related point concerns the (restricted) reflexivity of the logics proposed here, which again runs counter to the intuition of induction as generalisation from instances to populations. We expect names of individuals to be present in the inductive premisses, and absent from the inductive hypothesis. As a consequence, changing the names in the premisses (in a way that does not distort the information they convey) would not invalidate the hypothesis. For instance, consider the observations

$$\text{Raven}(a) \wedge \text{Black}(a) \wedge \neg \text{Raven}(c) \wedge \text{Black}(c) \wedge \neg \text{Raven}(d) \wedge \neg \text{Black}(d)$$

Given these observations, $\forall x : \text{Raven}(x) \rightarrow \text{Black}(x)$ is a possible hypothesis. In the confirmatory setting this can be formalised by constructing a regular model from the evidence, and stipulating that any hypothesis be satisfied by that model. However, from that model we can construct another model by replacing a, c, d with (say) e, f, g . It makes sense to say that this constitutes another regular model that is to satisfy any hypothesis — this would rule out any hypothesis that talks about a, c or d , including the observations. In terms of the confirmatory structures of Section 5.2 this would mean to drop the condition that $[\alpha] \subseteq \llbracket \alpha \rrbracket$ for all $\alpha \in L$. In this way the perception of induction as reasoning from the particular to the

²¹This may require a substantial background theory, limiting the practical feasibility of this encoding.

general would be realised, not by posing syntactic restrictions (which is logically unattractive), but by wiring it into the semantics.

6.2 *Logic*

In my view, a major concern of logic is the systematic study of *reasoning forms* (deduction, induction, plausible reasoning, counterfactual reasoning, and so on). The emphasis on correct or mathematical reasoning is much more recent:

The central process of reasoning studied by modern logicians is the accumulative *deduction*, usually explained semantically, as taking us from truths to further truths. But actually, this emphasis is the result of a historical contraction of the agenda for the field. Up to the 1930s, many logic textbooks still treated deduction, induction, confirmation, and various further forms of reasoning in a broader sense as part of the logical core curriculum. And moving back to the 19th century, authors like Mill or Peirce included various non-deductive modes of reasoning (induction, abduction) on a par with material that we would recognize at once as ‘modern’ concerns. Since these non-deductive styles of reasoning seemed irrelevant to foundational research in mathematics, they moved out quietly in the Golden Age of mathematical logic. But they do remain central to a logical understanding of ordinary human cognition. These days, this older broader agenda is coming back to life, mostly under the influence of Artificial Intelligence, but now pursued by more sophisticated techniques – made available, incidentally, by advances in mathematical logic...

vB

From a logical point of view this chapter can be seen as a contribution to the systematic study of reasoning forms, by trying to nail down the essence of induction in logical terms. By employing consequence relations as the central notion in this analysis, rather than, say, introducing non-truth-functional connectives or modalities, attention is focused on the underlying inference mechanism. Abstract properties of consequence relations can be used to classify reasoning forms: for instance, we could say that a reasoning form is deductive if it satisfies Monotonicity, quasi-deductive if it satisfies Cautious Monotonicity, explanatory if it satisfies (Admissible) Converse Entailment, and so on. These properties can be used to chart the whole map of reasoning forms, just like the properties of [Kraus *et al.*, 1990] chart the map of plausible reasoning. Thus, the present chapter can be seen as a constructive proof of the thesis that the method of analysis through consequence relations, pioneered by Gabbay [1985] Makinson [1989], and Kraus, Lehmann and Magidor [1990; 1992], constitutes in fact a *methodology*, that can be applied to analyse arbitrary forms of reasoning.

It is important to note that, by pursuing the analysis on the level of consequence relations, one is studying a class of logics, i.e. a reasoning form, rather than a

particular logic. The formal analysis of a reasoning form is quite different from the material definition of a particular logic. The traditional picture of the latter process is like this. One starts with the semantics, which is designed to provide a precise meaning for the primitive symbols in the language, and formalises the relevant notion of consequence. Only then the proof-theoretic axiomatisation follows, accompanied by proofs of soundness and completeness. However, the abstract analysis of a class of logics may well proceed along different lines. Usually, a number of material definitions of specific logics are available (e.g. default logic, circumscription, negation as failure), and one tries to understand what these logics have in common. This extraction of commonalities may start on the semantic level (e.g. each of the logics selects among the models of the premisses) but also on the meta-theoretical level (e.g. each of the logics is nonmonotonic, and closed under conjunction on the right-hand side). Semantics does not necessarily come first anymore: the notion of Cautious Monotonicity may add as much to the understanding of plausible reasoning as the idea of a preference ordering on models.

Another tendency that can be observed when moving towards more abstract characterisations of reasoning forms is that semantics concentrates more on a characterisation of the notion of consequence, and less on the meaning of the primitive symbols in the language. This raises the question as to what constitutes a logical semantics. This is by no means a settled issue, but different authors have put forward the notion of *preservation* as playing a central role in semantics. For instance, 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].

These points underline that many open problems of contemporary logic are conceptual in nature, rather than just technical. Motivated by problems from artificial intelligence, logic is widening its scope (again) to include forms of reasoning that are less and less similar to classical deduction. This development is far from nearing its completion. For instance, it seems commonplace among researchers studying consequence relations to assume that any consequence relation should minimally satisfy Reflexivity, Right Weakening, and Right And. Surely this makes sense if one limits attention to quasi-deductive reasoning, but the ex-

planatory consequence relations studied in Section 4 of this chapter satisfy none of these properties. This chapter makes a case for a much more liberal perception of what is a consequence relation.

6.3 *Machine learning*

Whereas the main contributions of this chapter are logical and philosophical in nature, it also provides a novel perspective on inductive machine learning. First of all, the framework of inductive logic elaborated above provides a new logical foundation of inductive learning. As such, viewing inductive learning through logic is of course not new, as is witnessed by the subdiscipline of inductive logic programming (ILP) [Muggleton and De Raedt, 1994]. However, the usual logical perception of inductive learning differs from the one presented here. These two perceptions can be explained by considering the phrase ‘inductive logic programming’. This phrase is ambiguous: it can be taken to mean — as it is usually done — ‘doing logic programming inductively’, but it can also be parsed as *programming in inductive logic* — which corresponds to the perception elaborated in this chapter. Let me explain why these interpretations are fundamentally different.

By identifying ILP with doing logic programming inductively, one effectively says that one’s main goal is logic programming, i.e. answering queries by executing a declarative specification by means of its procedural semantics; however, since this declarative specification is only partly known through a number of examples, we should do some inductive patching before the real work can start. This results in a somewhat subsidiary view of induction as a subproblem that needs to be solved before we can do the main task. This can be seen from the problem specification (see Section 2.4), which defines induction as a sort of reversed deduction from positive examples p to hypothesis H . The slogan *ILP = inductive logic + programming* offers a different viewpoint, from which the inference from examples to a logic program is the main step. That is, the examples (and background theory) provide the declarative specification of the induction task, which is executed by applying the inference rules of an inductive logic (e.g. specialisation or generalisation operators). The hypothesis is an inductive consequence of the examples.

An immediate advantage of the latter viewpoint is that it provides an independent definition of induction, instead of defining it in terms of something else (e.g. reversed deduction). The following analogy may clarify this point. In mathematics, many concepts are introduced as inverses of other concepts: division as inverse of multiplication, roots as inverses of powers, integration as inverse of differentiation, and so on. However, once such a concept has been introduced in this way, it usually gets an independent treatment, providing further insight in and justification of the new concept. For instance, the definition of a definite integral as the limit of a Riemann sum formalises the idea that a definite integral calculates the area under a curve. The relationship between the new concept and previously defined concepts is then obtained as a theorem (cf. the fundamental theorem of

calculus), rather than a seemingly arbitrary definition.

The framework of inductive consequence relations can be used to obtain soundness and completeness proofs of particular sets of operators employed by a particular induction algorithm. As an illustration, consider the inverse resolution operator absorption [Muggleton and Buntine, 1988]: from $p \leftarrow q, r$ and $s \leftarrow q, r, t$ infer $s \leftarrow p, t$. In our framework, a soundness proof of this induction step is established by the derivation, from an appropriate rule system such as **EM**, of the following statement:

$$\frac{\vDash p \leftarrow q, r}{s \leftarrow q, r, t \mid s \leftarrow p, t}$$

which should be read as follows: if $p \leftarrow q, r$ is known from the background knowledge, then $s \leftarrow p, t$ is an inductive consequence of $s \leftarrow q, r, t$.

The framework also provides a well-defined *vocabulary* for reasoning about induction tasks and algorithms. Important notions like incrementality and convergence are linked to the underlying inductive consequence relation. This is important because, as we have seen, different induction tasks may have different characteristics — articulating these characteristics is a necessary and important first step in understanding the induction task at hand, and choosing or devising the right algorithm. Using such a vocabulary we can construct a taxonomy of induction tasks. Two families of induction tasks have been discerned in this chapter: explanatory induction, aimed at obtaining classification rules, and confirmatory induction, directed towards extracting structural regularities from the data. While explanatory induction corresponds to typical classification-oriented inductive machine learning tasks such as concept learning from examples, induction tasks belonging to the non-classificatory paradigm have only recently started to attract attention [Helft, 1989; Flach, 1990; De Raedt and Bruynooghe, 1993; Flach and Lachiche, forthcoming]. The contribution of the present chapter has been to propose Hempel’s notion of qualitative confirmation as the unifying concept underlying these latter approaches. As Helft observed, inductive conclusions may be obtained by closed-world reasoning if the evidence may be considered complete; however, above I have proposed to distinguish an open variant of confirmatory reasoning, where the completeness assumptions on the evidence are considerably relaxed. While this may lead to an increase in the set of inductive hypotheses not refuted by given evidence, it has the distinct advantage of invalidating the rather strong property of Right And, and restoring the computationally attractive property of Convergence.

7 CONCLUSIONS

This chapter has been written in an attempt to increase our understanding of inductive reasoning through logical analysis. What logic can achieve for arbitrary forms of reasoning is no more and no less than a precise definition of what counts as a *possible* conclusion given certain premisses. Selecting the best (most useful,

most plausible, etc.) hypothesis is an extra-logical matter. The logic of induction is the logic of inductive hypothesis generation.

There is not a single logic of induction. The logical relationship between evidence and possible inductive hypotheses depends on the task these hypotheses are intended to perform. Induction of classification rules such as concept definitions are based on a notion of explanation; an alternative logical account of induction starts from the qualitative relation of confirmation. Other forms of induction are conceivable. In this chapter I have proposed a meta-level framework for characterising and reasoning about different forms of induction. The framework does not fix a material definition of inductive hypothesis generation, but can be used to aggregate knowledge about classes of such material logics of induction.

A number of technical results have been obtained. The system **EM** axiomatises explanation-preserving reasoning with respect to a monotonic explanation mechanism. Characterisation of explanatory induction with respect to weaker (e.g. preferential) explanation mechanisms is left as an open problem. The systems **CS** and **CP** axiomatise the general and preferential forms of closed confirmatory reasoning, conceived as reasoning about selected regular models. They represent variations of earlier, differently motivated, characterisations by Bell [1991] and Kraus, Lehmann and Magidor [1990], with a different treatment of inconsistent premisses. An important open problem here is the axiomatisation of confirmatory structures where regular models may not be models of the premisses. Finally, the system **CW** represents an extreme form of open confirmatory reasoning (i.e. compatibility of premisses and hypothesis). Finding more realistic forms of open confirmatory reasoning remains an open problem.

ACKNOWLEDGEMENTS

I like to thank John-Jules Meyer and Daniel Lehmann for stimulating discussions. Part of this work was supported by Esprit IV Long Term Research Project 20237 (Inductive Logic Programming 2), and by the EU-funded Networks of Excellence ILPnet and ILPnet2.

REFERENCES

- [Angluin and Smith, 1983] D. Angluin and C. H. Smith. Inductive inference: theory and methods. *Computing Surveys*, **15**, 238–269, 1983.
- [Bell, 1991] J. Bell. Pragmatic logics. In *Proc. Second International Conference on Knowledge Representation and Reasoning KR'91*, pp. 50–60, Morgan Kaufmann, San Mateo, 1991.
- [Van Benthem, 2000] J. van Benthem. Reasoning in reverse. In [Flach and Kakas, 2000].
- [Carnap, 1950] R. Carnap. *Logical Foundations of Probability*. Routledge & Kegan Paul, London, 1950.
- [Delgrande, 1987] J. P. Delgrande. A formal approach to learning from examples. *Int. J. Man-Machine Studies*, **26**, 123–141, 1987.
- [De Raedt and Bruynooghe, 1993] L. De Raedt and M. Bruynooghe. A theory of clausal discovery. In *Proc. 13th International Conference on Artificial Intelligence IJCAI'93*, pp. 1058–1063. Morgan Kaufmann, San Mateo, 1993.

- [Flach, 1990] P. A. Flach. Inductive characterisation of database relations. In *Proc. International Symposium on Methodologies for Intelligent Systems ISMIS'90*, Z.W. Ras, M. Zemankowa and M. L. Emrich (eds.), pp. 371-378. North-Holland, Amsterdam.
- [Flach, 1993] P. A. Flach. Predicate invention in Inductive Data Engineering. In *Proc. European Conference on Machine Learning ECML'93*, P. B. Brazdil, ed. pp. 83-94. Lecture Notes in Artificial Intelligence 667, Springer-Verlag, Berlin, 1993.
- [Flach, 1995] P. A. Flach. *Conjectures: an inquiry concerning the logic of induction*. PhD thesis, Tilburg University, 1995.
- [Flach, 1996] P. A. Flach. Rationality postulates for induction. In *Proc. Theoretical Aspects of Rationality and Knowledge TARK'96*, Yoav Shoham, ed. pp. 267-281. Morgan Kaufmann, San Mateo, 1996.
- [Flach, 1998] P. A. Flach. Comparing consequence relations. In *Proc. Sixth International Conference on Knowledge Representation and Reasoning KR'98*, A. G. Cohn, L. Schubert and S. C. Shapiro, eds. pp. 180-189. Morgan Kaufmann, San Mateo, 1998.
- [Flach and Savnik, 1999] P. A. Flach and I. Savnik. Database dependency discovery: a Machine Learning approach. *AI Communications* **12**(3), 139-160, 1999.
- [Flach and Lachiche, forthcoming] P. A. Flach and N. Lachiche. Confirmation-guided discovery of first-order rules with Tertius. *Machine Learning*, accepted for publication.
- [Flach and Kakas, 2000] P. A. Flach and A. C. Kakas, eds. *Abductive and inductive reasoning: essays on their relation and integration*. Kluwer Academic Publishers, 2000.
- [Gabbay, 1985] D. M. Gabbay. Theoretical foundations for non-monotonic reasoning in expert systems. In *Logics and Models of Concurrent Systems*, K. R. Apt, ed. pp. 439-457. Springer-Verlag, Berlin, 1985.
- [Hanson, 1958] N. R. Hanson. The logic of discovery. *J. Philosophy*, **55**, 1073-1089, 1958.
- [Helft, 1989] N. Helft. Induction as nonmonotonic inference. In *Proc. First International Conference on Knowledge Representation and Reasoning KR'89*, pp. 149-156. Morgan Kaufmann, San Mateo, 1989.
- [Hempel, 1943] C. G. Hempel. A purely syntactical definition of confirmation. *J. Symbolic Logic*, **6**, 122-143, 1943.
- [Hempel, 1945] C. G. Hempel. Studies in the logic of confirmation. *Mind*, **54**, 1-26 (Part I); **54**, 97-121 (Part II), 1945.
- [Jennings *et al.*, 1991] R. E. Jennings, C. W. Chan and M. J. Dowad. Generalised inference and inferential modelling. In *Proc. Int. Joint Conf. on Artificial Intelligence IJCAI'91*, pp. 1046-1051. Morgan Kaufmann, 1991.
- [Kraus *et al.*, 1990] S. Kraus, D. Lehmann and M. Magidor. Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, **44**, 167-207, 1990.
- [Lachiche, 2000] N. Lachiche. Abduction and induction from a non-monotonic reasoning perspective. In [Flach and Kakas, 2000].
- [Lehmann and Magidor, 1992] D. Lehmann and M. Magidor. What does a conditional knowledge base entail? *Artificial Intelligence*, **55**, 1-60, 1992.
- [Makinson, 1989] D. Makinson. General theory of cumulative inference. In *Proc. 2nd International Workshop on Non-Monotonic Reasoning*, M. Reinfrank, J. de Kleer, M. L. Ginsberg and E. Sandewall, eds. pp. 1-18. Lecture Notes in Artificial Intelligence 346, Springer-Verlag, Berlin, 1989.
- [Mill, 1843] J. S. Mill. *A System of Logic*, 1943. Reprinted in *The Collected Works of John Stuart Mill*, J. M. Robson (ed.), Routledge and Kegan Paul, London.
- [Muggleton and De Raedt, 1994] S. Muggleton and L. De Raedt. Inductive Logic Programming: theory and methods. *J. Logic Programming* **19-20**, 629-679, 1994.
- [Muggleton and Buntine, 1988] S. Muggleton and W. Buntine. Machine invention of first-order predicates by inverting resolution. In *Proc. Fifth International Conference on Machine Learning*, J. Laird, eds. pp. 339-352. Morgan Kaufmann, San Mateo, 1988.
- [Tarski, 1936] A. Tarski. Über den Begriff der logischen Folgerung. *Actes du Congrès Int. de Philosophie Scientifique*, **7**, 1-11, 1936. Translated into English as On the concept of logical consequence. In *Logic, Semantics, Metamathematics*, A. Tarski, pp. 409-420. Clarendon Press, Oxford, 1956.
- [Zadrozny, 1991] W. Zadrozny. On rules of abduction. IBM Research Report, August 1991.