On the logic of induction

Peter A. Flach[*]

INFOLAB, Tilburg University, PObox 90153, 5000 LE Tilburg, the Netherlands

Abstract.

This paper 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 *formation* rather than evaluation. Inspired by the seminal paper of Kraus, Lehmann & Magidor [18] we analyse the logic of inductive hypothesis formation on the metalevel 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 paper is extensively motivated by referring to recent and older work in philosophy, logic, and Machine Learning.

1. Introduction

This paper is an attempt 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. The main contribution of this paper lies in the novel perspective that is obtained 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.

1.1 Problem formulation and approach

I should start by stressing that the study reported on in this paper should be perceived as an application of logical analysis to problems in Artificial Intelligence. Thus, 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

[*] E-mail: Peter.Flach@kub.nl.

controversies abounding in philosophy of science as to whether or not science proceeds by inductive methods. This is not to say that I will completely ignore philosophical considerations — 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 paper 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 induction’ I mean the by now standard approach, developed in the first half of this century, 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 paper.

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

The objective pursued in this paper, then, is to develop semantics and proof systems for inductive hypothesis formation. What is new here is not so much this objective, which has been pursued before (see e.g. [4]), but the meta-theoretical viewpoint taken in this paper, 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 & Magidor [18], 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 paper as a constructive proof of the thesis that their techniques in fact establish a *methodology*, by demonstrating how they can be successfully applied to analyse a rather different form of reasoning.

1.2 Plan of the paper

The paper is structured as follows. In section **2** the philosophical, logical, and Machine Learning backgrounds of this paper are surveyed. Section **3** introduces the main logical tool employed in this paper: the notion of a metalevel consequence relation. Sections **4** and **5** form the technical core of this paper, stating representation theorems characterising sets of metalevel properties of explanatory induction and confirmatory induction, respectively. In section **6** we discuss the implications of the approach taken and results obtained in this paper. Section **7** repeats the main conclusions.

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 socalled 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 [3]. 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.' [3, §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 paper, 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.

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

[1] Note that degree of confirmation is not a quantity that is simply to be maximised, since this would lead us straight back into deductive logic.

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 the same truth-estimating procedure for very different forms of reasoning.

2.2 Confirmation as a qualitative relation

Carl G. Hempel [15, 16] 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 acount. 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 metalevel axioms, and we will discuss them at some length. The following conditions can be found in [16, 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' [16, 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 [16, 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'. We will return to material definitions of qualitative confirmation in section *2.5*.

2.3 Abduction

Predating 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, which I will briefly discuss in this section.

In a series of lectures on Pragmatism delivered in 1903, Peirce distinguishes three types of reasoning: deduction, induction, and abduction. Induction 'consists in starting from a theory, deducing from it predictions of phenomena, and observing those phenomena in order to see *how nearly* they agree with the theory'. Furthermore,

> 'The justification for believing that an experiental theory which has been subjected to a number of experimental tests will be in the near future sustained about as well by further such tests as it has hitherto been, is that by steadily pursuing that method we must in the long run find out how the matter really stands.' [13, 5.170]

Note that Peirce claims, like Carnap, that induction evaluates the plausibility of a given theory, rather than constructing that theory from observations. However, inductive hypotheses do not come out of the blue, and this is where abduction comes into play:

> 'Abduction is the process of forming an explanatory hypothesis. It is the only logical operation which introduces any new idea; for induction does nothing but determine a value, and deduction merely evolves the necessary consequences of a pure hypothesis.
>
> Deduction proves that something *must* be; Induction shows that something *actually is* operative; Abduction merely suggests that something *may be*.
>
> Its only justification is that from its suggestion deduction can draw a prediction which can be tested by induction, and that, if we are ever to learn anything or to understand phenomena at all, it must be by abduction that this is to be brought about.
>
> No reason whatsoever can be given for it, as far as I can discover; and it needs no reason, since it merely offers suggestions.' [13, 5.171]

In other words, *abduction is the process of conjecturing inductive hypotheses*, constrained by the requirement that they should comply with the available observations. Abduction represents the purely logical part of inductive reasoning.[2]

Peirce proceeds by defining the logical form of abduction. 'It must be remembered', he writes, 'that abduction, although it is very little hampered by logical rules, nevertheless is logical inference, asserting its conclusion only problematically or conjecturally, it is true, but nevertheless having a perfectly definite logical form.' Peirce then defines this logical form, as follows.

[2]Unfortunately, the term 'abduction' is nowadays used in two different ways. Peirce himself is to blame at least partly for this confusion, sine he first proposed a rather different, syllogistic classification of reasoning forms, which can be summarized as follows. Consider the Aristotelian syllogism *Barbara*: 'All the beans from this bag are white; these beans are from this bag; therefore, these beans are white'. Now there are two ways to exchange the conclusion with one of the premisses, one resulting in the inductive syllogism 'These beans are white; these beans are from this bag; therefore, all the beans from this bag are white', the other in 'All the beans from this bag are white; these beans are white; therefore, these beans are from this bag'. Peirce refers to this latter syllogism as (forming a) *hypothesis.* This syllogistic theory has to a large extent been adopted in the discipline of logic programming, where abduction (ironically, the term was only introduced in Peirce's later theory) is generally perceived as the inference of ground facts from rules and a query that is to be explained. Notice that a logic based on entailment rather than syllogisms is unable to distinguish between the two latter syllogisms, which both embody a form of reversed deduction. See [10].

> 'Long before I first classed abduction as an inference it was recognized by logicians that the operation of adopting an explanatory hypothesis — which is just what abduction is — was subject to certain conditions. Namely, the hypothesis cannot be admitted, even as a hypothesis, unless it be supposed that it would account for the facts or some of them. The form of inference, therefore, is this:
>
> The surprising fact, C, is observed;
>
> But if A were true, C would be a matter of course,
>
> Hence, there is reason to suspect that A is true.
>
> Thus, A cannot be abductively inferred, or if you prefer the expression, cannot be abductively conjectured until its entire content is already present in the premiss, "If A were true, C would be a matter of course." '
>
> [13, 5.188]

In short, the view of induction that Peirce offers here [redacted] this. Inductive reasoning consists of two steps: (*i*) formulating a conjecture, and (*ii*) evaluating the conj[redacted]ture. Both steps take the available evidence into account, but in quite different ways and with different [redacted]als. The first step requires that the conjectured hypothesis *explains* the observations; having a definite l[redacted]ical form, it represents a form of inference. The second step evaluates how well predictions offered [redacted] reality; it is not inference, but assigns a numerical value to a hypothes[redacted]n order to avoid terminological problems, I will not use Peirce's terminology and refer to the first st[redacted] explanatory hypothesis formation, and to the second as hypothesis evaluation or validation.

Leaving a few details aside, Peirce's definition of [redacted]natory hypothesis formation can be formalised as the inference rule

$$\frac{C\ ,\ \text{[redacted]}\ C}{\text{[redacted]}}$$

In this paper I propose to generalise Peirce's definitio[redacted] by including the relation of 'is explained by' as a parameter. This is achieved by lifting the explanatory [redacted]ference from C to A to the metalevel, as follows:

$$\frac{A\ \text{[redacted]}\ C}{C\ \text{[redacted]}\ A}$$

The symbol $\mid\!<$ stands for the explanatory consequence relation. Axiom systems for this relation will be considered in section **4**.

2.4 Confirmation vs. explanation

We now have encountered two fundamental notions that play a role in inductive hypothesis formation: one is that the hypothesis should be *confirmed* by the evidence, the other that the hypothesis should *explain* the evidence. 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 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?' [16, p.104]. 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 paper I defend the position that Hempel was reluctant to take, namely that with respect to inductive or scientific hypothesis formation 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.

2.5 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, in which instances and concepts are described, 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 [redacted]efer to instances and non-instances. In this way, a classification of an instance can be represented [redacted] a truthvalue, which can be obtained by setting up a proof.[3] We then obtain the following general prob[redacted]m definition:

Problem: Concept learning from example[redacted]
Given: (1) A predicate-logical lang[redacted]
(2) A predicate representing [redacted]e target concept.
(3) Two sets P and N of gro[redacted]epresenting the positive and negative exa[redacted]
(4) A background theory T [redacted]nces.
Determine: A hypothesis H within th[redacted]provided language such that
(*i*) for all $p \in P$: $T \cup H$ [redacted]p;
(*ii*) for all $n \in N$: $T \cup H$ [redacted]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) [22]. 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 [6, 7, 8]. Furthermore, De Raedt & Bruynooghe have generalized the classificatory ILP-setting in order to induce non-Horn clauses from ground facts [5]. Both approaches essentially employ the following problem statement.

[3] The alternative is to represent concepts by open formulae, and to operationalize classification by means of subsumption.

Problem: Non-classificatory induction.
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 & Bruynooghe and myself is the truth-minimal Herbrand model of the evidence.[4] 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 dabases, 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”.’ [14, p.149]

There is a direct connection between Peirce’s conception of abduction as formation 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. In this paper 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.

3. Inductive consequence relations

In the sections to follow I employ the notion of a consequence relation, originating from Tarski [24] and further elaborated by Gabbay [11], Makinson [20], and Kraus, Lehmann & Magidor [18, 19]. In this section I give an introduction to this important metalogical 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 paper 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 metalanguage in which statements about the reasoning behaviour of the

[4]An alternative approach is to consider the information-minimal partial model of the evidence [8].

inductive agent are expresse[redacted] this [redacted]aper L [redacted]a propositional language[5] over a fixed coun[redacted] proposition symbols, closed u[redacted]er the[redacted]sual log[redacted]al connectives. We assume a set of propositio[redacted]al models U, and a satisfaction relation [redacted] with respect to the logical conn[redacted]tives and compact. As usual, we write [redacted] for ∀[redacted]∈U: m[redacted], for arbitrary $\alpha \in L$. Note that U may be a pr[redacted] of the set of all truth-assignm[redacted]ts to [redacted]oposition[redacted] symbols in L, which would reflect prior kn[redacted]ledge or *background knowledge* of the inducti[redacted]agent. [redacted]quivalently, we may think of U as the set of m[redacted]dels of an implicit background theory T, and let [redacted] stand[redacted]or 'α is a logical consequence of T'.

The metalanguage is a restricted p[redacted]nguage built up from a unary metapredicate [redacted]in prefix notation (standing for validity with respect [redacted]U in L) and a binary metapredicate $|\!\sim$ in inf[redacted]notation (standing for inductive consequence). In referr[redacted] to object-level formulae from L we employ a countable set of metavariables α, β, γ, δ, …, the logica[redacted] like function symbols on the metalevel), and the metaconstants **true** and **f**[redacted]**e**. Formulae of the metalanguage, usually referred to as *rules* or *properties*, are of the form $P_1, \ldots,$ [redacted]$/\ Q$ for $n \geq 0$, where $P_1, \ldots, P_n$ and Q are literals (atomic formulae or their negation). Intuitively, su[redacted]eted as an implication with *antecedent* $P_1, \ldots, P_n$ (interpreted conjunctiv[redacted]) and *consequent* Q, in which all variables are implicitly universally quantified. An example of such a r[redacted]e, written in an expanded Gentzen-style notation, is

[redacted]
$$\frac{\alpha\wedge\beta\rightarrow\gamma\ ,\ \alpha \mid\!\sim \beta}{\alpha\wedge\neg\gamma \not\mid\!\sim \beta}$$

This is a rule with t[redacted] positive literals in its antecedent, and a negative literal in its consequent. Intuitively, it express[redacted]that an inductive hypothesis β, previously inferred from evidence α, should be withdrawn if the nega[redacted]n of a consequence of α and β together is added to the evidence.

Consequence relat[redacted]*ns* provide the semantics for this metalanguage, by fixing the meaning of the metapredicate $|\!\sim$. Form[redacted]s a subset of $L \times L$. They will be used to model part or all of the reasoning be[redacted]viour of a particular reasoning agent, by listing a number of *arguments* (pairs of *premiss* and *conclusi*[redacted] accept[6]. A consequence relation satisfies a rule whenever it satisfies [redacted] violates it otherwise, where an instance of a rule is obtained by replacing [redacted] formulae from L. A consequence relation satisfies an instance of a rule if, [redacted]ls in the antecedent of the rule, it also satisfies the consequent. Finally:

- (*i*) a literal [redacted] is satisfied whenever the propositional formula from L denoted by α is true in e[redacted]ry model in U;
- (*ii*) a literal [redacted]α is satisfied whenever the propositional formula from L denoted by α is false in [redacted]me model in U;
- (*iii*) a literal $\alpha \mid\!\sim \beta$ is satisfied whenever the pair of propositional formulae from L denoted by α and β is an element of the consequence relation;
- (*iv*) a literal $\alpha \not\mid\!\sim \beta$ is satisfied whenever the pair of propositional formulae from L denoted by α and β is not 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. We will often refer to a particular consequence relation as $|\!\sim$ and write $p \mid\!\sim q$ instead of $\langle p,q\rangle \in |\!\sim$.

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 interpretations[7] of the metalanguage, whose rules can be easily transformed to clausal notation. The following terminology is borrowed from logic programming:

- (*i*) if all of $P_1,\ldots,P_n$ and Q are positive literals the r[redacted] is *definite*;
- (*ii*) if at least one of $P_1,\ldots,P_n$ is a negative literal an[redacted] is a positive literal the rule is *indefinite*;
- (*iii*) if all of $P_1,\ldots,P_n$ are positive literals and Q is a n[redacted]gative literal the rule is a *denial*.[8]

---

[5] One may argue that in induction the distinction between statements a[redacted]ut individuals (ground facts) and statements about sets of individuals (universal sentences) is crucial, which calls f[redacted]a predicate language. This point is discussed in section **6**.

[6] This is not to say that the agent actually draws the conclusion when o[redacted]erving the premisses, but merely that she considers it one of the possible conclusions.

[7] These Herbrand interpretations are restricted to the metapredicate $|\!\sim$; [redacted]s treated as a built-in predicate.

[8] This exhausts all the possibilities: the case that at least one of $P_1,\ldots,$[redacted] is a negative literal and Q is a negative

For instance, all the rule systems of [18] 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 **R** characterised in [19] contai[redacted] one indefinite rule, *viz.* Rational Monotonicity. Consequently, the set of rational consequence relations [redacted] not closed under intersection.

3.2 Some properties of inductive consequence relations

After having stated the main definitions concerning consequence relat[redacted]s I will now list some properties generally obeyed by inductive consequence relations. In this sectio[redacted]ween explanatory or confirmatory induction, and simply interpret $\alpha \mid\!\approx \beta$ as[redacted]given evidence α'.

The first two rules state that the logical form of evidence and hypoth[redacted]

Left Logical Equivalence [redacted]$\leftrightarrow\beta$, $\alpha \mid\!\approx \gamma$
$\beta \mid\!\approx \gamma$

Right Logical Equivalence [redacted]$\leftrightarrow\gamma$, $\alpha \mid\!\approx \beta$
[redacted]

Left Logical Equivalence states, for instance, that if the evidence is [redacted]pressed as a conjunction of ground facts, the order in which they occur is immaterial. From the viewpo[redacted]rithms this may embody a considerable simplification — however, the fra[redacted]aper is intended to provide a model for inductive reasoning in general rather [redacted]an particular algorithms.[9]

The following two rules express principles well-know[redacted] from philo[redacted]

Verification [redacted]$\wedge\beta\rightarrow\gamma$, $\alpha \mid\!\approx \beta$
$\alpha\wedge\gamma \mid\!\approx \beta$

Falsification [redacted]β
$\alpha\wedge\neg\gamma \not\mid\!\approx \beta$

In these two rules γ is a *prediction* made on the bas[redacted]dence α. Verification expresses that if such a prediction is indeed observed, [redacted]le hypothesis, while if its negation is observed, β may be considered refute[redacted]according to F[redacted]sification.[10] One might remark that typically the inductive hypothesis will entail the evi[redacted]nce, so that the[redacted]rst condition in the antecedent of Verification and Falsification may be simpl[redacted]ed to [redacted]$\rightarrow\gamma$[redacted]However, [redacted]his is only the case for certain approaches to explanatory induction; generally [redacted]eaking indu[redacted]ve hypot[redacted]ses, in pa[redacted]cular those that are confirmed without being an explanation, may [redacted]t contain all [redacted]e inform[redacted]on conve[redacted]d by the evidence. The formulation above represents the general c[redacted]e.

Falsification can be simplified in another sen[redacted]

LEMMA 3.1. *In the presence of Left Lo*[redacted]*cal Equivale*[redacted]*e, Falsif*[redacted] *the following rule:*

Consistency [redacted]$\alpha \mid\!\approx \beta$
[redacted]$\beta\rightarrow\neg\alpha$

Proof. To derive Falsification, suppose [redacted]$\wedge\beta\rightarrow\gamma$, i.e. [redacted]$\beta\rightarrow\neg(\alpha\wedge\neg\gamma)$, then b[redacted]onsistency $\alpha\wedge\neg\gamma \not\mid\!\approx \beta$. To derive Consistency fro[redacted]Falsification, [redacted]suppose $\alpha \mid\!\approx \beta$ and [redacted]$\beta\rightarrow\neg\alpha$, i.e.

---

literal can be rewritten to (*i*) or (*ii*).

[9]Practical 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).

[10]Notice 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 [1] the algorithm is conservative), while Falsification expresses that the current hypothesis must be abandoned when the next observations runs counter to the predictions of the algorithm (called consistency by [1]). However, in the context of the present paper these are not the intended interpretations of the two rules.

[redacted]$\wedge\beta\rightarrow$**false**, then by Falsification $\alpha\wedge\neg$**false**$\not\mathrel{|\!\!<}\beta$, and by Left Logical Equivalence $\alpha\not\mathrel{|\!\!<}\beta$, [redacted] 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. 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:[11]

Left Reflexivity [redacted] $\dfrac{\alpha\mathrel{|\!\!<}\beta}{\alpha\mathrel{|\!\!<}\alpha}$

Right Reflexivity $\dfrac{\alpha\mathrel{|\!\!<}\beta}{\text{[redacted]}}$

If a consequence relation contains an argument $\alpha\mathrel{|\!\!<}\alpha$, this signals th[redacted] α is consistent with the reasoner's background theory. We will call such an α *admissible* (with respect [redacted] the consequence relation), and use conditions of this form whenever we require consistency of evidence [redacted]

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

Right Extension [redacted] $\dfrac{\wedge\beta\rightarrow\gamma\ ,\ \alpha\mathrel{|\!\!<}\beta}{\alpha\mathrel{|\!\!<}\beta\wedge\gamma}$

Further rules considered in this paper 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 moving to the technical results of the paper 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 [redacted] reasoning agent. The symbol $\mathrel{|\!\!<}$ is introduced in order to reason about consequence relations and [redacted]asoning behaviour, and functions as a binary predicate in the metalanguage. Rules describing pro[redacted]rties of $\mathrel{|\!\!<}$ express boundaries of rationality of inductive reasoning: a consequence relation violating such [redacted] rule would be considered irrational.

Now suppo[redacted] **X** is a set of such rationality postulates, and let *A* be a set of inductive arguments. Clearly, by m[redacted]itional inductive arguments *A′*, the significance of which is: if an inductive a[redacted]nt accepts arguments *A*, and it behaves rationally according to **X**, it should also accept arguments *A′*. [redacted]or instance, if **X** includes the rule of Verification, *A* contains the argument

[redacted]ad_is_black $\mathrel{|\!\!<}$ crows_are_black

and the backg[redacted]und knowledge includes

[redacted]rows_are_black→chevy_is_black

then *A′* contains the additional inductive argument

---

[11] One 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**.

[12] Readers with a background in inductive learning may interpret $\alpha\mathrel{|\!\!<}\alpha$ as 'hypothesis α does not cover any negative example'.

$$\texttt{chad\_is\_black} \wedge \texttt{chevy\_is\_black} \mid\!\!< \texttt{crows\_are\_black}$$

implying that if the agent would not accept the latter, it would behave irrationally (wrt. **X**).

One may now ask: what is the smallest set of arguments containing A and satisfying the rules of **X**? Such a set, if it exists, would represent the *closure* of A under **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 **X**. Clearly, the existence of such a closure operation depends on the rules in **X**. As has been remarked before, if all the rules in **X** are definite (having only positive literals in their antecedents and consequent) the closure of A under **X** is unique for arbitrary A. All the rule systems in [18] 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 [19]; other examples will be found in this paper. 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 **X**, and we take their intersection to find out on what arguments they agree, this intersection itself may not satisfy the rules of **X**.

Lehmann & Magidor are not content with this indefiniteness: they define an operation of *rational closure* which selects, among the many supersets of A that satisfy **X**, one that has certain desirable properties [19, p.33]. This seems to be motivated by the tacit assumption that *rationality postulates* **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 [24], 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 metalevel, closure operations establish a *meta-metalevel*.

In this paper we will not be concerned with inference on this meta-metalevel. 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-metalevel view is that it obscures the status of the consequence relation symbol. The relevant question on the meta-metalevel is: what expressions of the form $\alpha \mid\!\sim \beta$ are entailed by a set of expressions of the same form? In other words: on the meta-metalevel $\mid\!\sim$ acts as a *connective.*[13] This is also apparent from Kraus, Lehmann & Magidor's terminology: they call $\alpha \mid\!\sim \beta$ 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-metalevel perspective is at odds with the metalevel 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.

In this paper the notion of closure will be employed on the metalevel rather than the meta-metalevel, in order to compare consequence relations and rule systems. Given a consequence relation $\mid\!\!<$, its *closure* $C_{\mid\!<}$: $L \rightarrow 2^L$ is defined as $C_{\mid\!<}(\alpha) = \{\beta \mid \alpha \mid\!\!< \beta\}$, and $C_{\mid\!<}(\alpha)$ is referred to as the closure of α under $\mid\!\!<$. A consequence relation $\mid\!\!<_1$ is called (at least) *as restrictive as* another consequence relation $\mid\!\!<_2$ if for every $\alpha \in L$ the closure of α under $\mid\!\!<_1$ is a subset of the closure of α under $\mid\!\!<_2$ — or equivalently[redacted] if [redacted] is a subset of $\mid\!\!<_2$ — and *more restrictive than* $\mid\!\!<_2$ if in addition $\mid\!\!<_1 \neq \mid\!\!<_2$. A set of rules $\mathbf{X_1}$ is (at leas[redacted]*as*[redacted]*strictive as* another set of rules $\mathbf{X_2}$ if for every consequence relation $\mid\!\!<_2$ satisfying $\mathbf{X_2}$ there is a uniq[redacted] le[redacted]t restrictive consequence relation $\mid\!\!<_1$ satisfying $\mathbf{X_1}$ such that $\mid\!\!<_1$ is as restrictive as $\mid\!\!<_2$; we say t[redacted]t [redacted] is the $\mathbf{X_1}$-*restriction* of $\mid\!\!<_2$. In addition the mapping from $\mid\!\!<_2$ to its $\mathbf{X_1}$-restriction is required to be [redacted] the set of relations satisfying $\mathbf{X_1}$ (see [9] for further motivation and analysis of these de[redacted]iti[redacted]s).

The preceding definitions reflect that rule systems should be compared by com[redacted]ri[redacted] the set of conclusions for given premisses, rather than by metalevel entailment. From [KLM90] [redacted]

[13]This raises the question why conditional assertions cannot be nested, as in $(\alpha \mid\!\sim \beta) \mid\!\sim \gamma$. Note tha[redacted]he [redacted]swer to this question is perfectly clear on the metalevel, since this expression makes as little sense as, say, $(\alpha$ [redacted] $\beta)$ [redacted] γ.

impression that monotonic consequence is stronger (more restrictive) than preferential consequence because the rule system **M** entails every rule in the rule system **P** (or equivalently, every monotonic consequence relation is preferential). However, this does not work in general: the metalevel axiom $\alpha \mathrel{|\!\sim} \beta$ entails all the rules in **M**, yet defines a very unrestrictive form of reasoning. Furthermore, our criterion also allows to compare rule systems that are not related by metalevel entailment, as we will see below.

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 $\alpha \mathrel{|\!<} \beta$ 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 [redacted] formulate abstract properties of hypothesis formation, without fixing a particular material definition. [redacted]e will then single out a particular set of properties (the system **EM**) and characterise it semanticall[redacted]by means of strong explanatory structures.

4.1 Properties of explanatory induction

A natural requirement for explanatory induction is that every c[redacted]ls the evidence counts as a possible explanation. As explained above[redacted]condition that hypothesis β be consistent is expressed by $\beta \mathrel{|\!<} \beta$, which gives us the following rule:

Admissible Converse Entailment [redacted] $\dfrac{\rightarrow\alpha \ , \beta \mathrel{|\!<} \beta}{\alpha \mathrel{|\!<} \beta}$

Another requirement for explanations has been discussed above as[redacted]H5′): possible explanations may be logically strengthened, as long as they remain consistent. This is ex[redacted]ssed as follows:

Admissible Right Strengthening $\dfrac{\alpha \mathrel{|\!<} \beta \ , \text{[redacted]} \rightarrow\beta \ , \gamma \mathrel{|\!<} \gamma}{\alpha \mathrel{|\!<} \gamma}$

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

Explanatory[redacted]eflexivity $\dfrac{\alpha \mathrel{|\!<} \alpha \ , \neg\beta \not\mathrel{|\!<} \alpha}{\beta \mathrel{|\!<} \beta}$

This rule represents a [redacted]ecially tailored for explanatory induction. It i[redacted] understood by rewritin[redacted] into its contrapositive: from $\alpha \mathrel{|\!<} \alpha$ and $\beta \not\mathrel{|\!<} \beta$ infer $\neg\beta \mathrel{|\!<} \alpha$, which states t[redacted] β is inadmissible, i.e. t[redacted]strong a statement with regard to the background knowledge, its negation[redacted] so weak that it is explai[redacted]ypotheses α.

> LEMMA 4.1. *In [redacted]e presence of Consistency and Explanatory Reflexivity, Admissible Right Strengthening i[redacted]lies Admissible Converse Entailment.*
> Proof. Suppose [redacted]$\rightarrow\alpha$, then by Consistency $\neg\alpha \not\mathrel{|\!<} \beta$. Suppose furthermore $\beta \mathrel{|\!<} \beta$, then by Explanatory R[redacted]exivity $\alpha \mathrel{|\!<} \alpha$. 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:[14]

[14]In previous work [8] Incrementality was called Additivity, and Convergence was called Incrementality. The

Incrementality [redacted] $\mid\!\sim \gamma , \beta \mid\!\sim \gamma$ / $\alpha\wedge\beta \mid\!\sim \gamma$

Convergence [redacted]$\rightarrow\beta$, $\alpha \mid\!\sim \gamma$ / $\beta \mid\!\sim \gamma$

LEMMA 4.2. *If* $\mid\!\sim$ *is a consequence relation satisfying Incrementality and Convergence, then* $\alpha\wedge\beta \mid\!\sim \gamma$ *iff* $\alpha \mid\!\sim \gamma$ *and* $\beta \mid\!\sim \gamma$,.
Proof. The *if* part is Incrementality, and the *only-if* part follows from Convergence.

Incrementality and Convergence are of considerable importance for computational induction, since they allow for an incre[redacted]ntal approach. Incrementality states that pieces of evidence can be dealt v[redacted] isolation. Another way [redacted]say the same thing is that the set of evidence explained by a given hypoth[redacted] conjunctively closed. [redacted] the rule of Consistency this set is consistent, which yields the foll[redacted] principle:

Left Consist[redacted]cy $\alpha \mid\!\sim \beta$ / $\neg\alpha \not\mid\!\sim \beta$

LEMMA 4.3. *In t*[redacted] *presence of Right Reflexivity and Admissible Converse Entailment, Left Consistency imp*[redacted]*s Consistency.*
Proof. Suppose [redacted]$\beta\rightarrow\neg\alpha$. Now, either $\beta \mid\!\sim \beta$ or $\beta \not\mid\!\sim \beta$; in the former case, $\neg\alpha \mid\!\sim \beta$ by Admissible Con[redacted]rse Entailment, and we conclude by Left Consistency. In the latter case, we have $\delta \not\mid\!\sim \beta$ for any δ by Right Reflexivity.

It follows that Left Consistency and Consistency are equivalent in[redacted]he presence of Right Reflexivity, Admissible Converse Entailment, and Incrementality.

Convergence states a *monotonicity* property of induction, whic[redacted]can again best be understood by considering its contrapositive: a hypothesis that is rejected on[redacted]he [redacted]sis of evidence β cannot become feasible again when stronger evidence α is available. In other w[redacted]ds: [redacted] is not defeasible (i.e. based on assumptions), but based on [redacted] monotonicity property of deduction (note that the latter c[redacted]be[redacted]ta[redacted]ed b[redacted]versing the implica[redacted] the first condition of Convergence).

LEMMA 4.4. *The combination of Verification an*[redacted] *following rule:*

Predictive Convergence [redacted]

Proof. To derive Predictive Convergence, sup[redacted]$\alpha\wedge\gamma\rightarrow\beta$ [redacted] $\alpha \mid\!\sim \gamma$, then by [redacted] Verification $\alpha\wedge\beta \mid\!\sim \gamma$, and by Convergence $\beta \mid\!\sim \gamma$.
Predictive Convergence implies Convergence, since[redacted]$\rightarrow\beta$ implies [redacted]$\gamma\rightarrow\beta$.
Predictive Convergence implies Verification, since [redacted]$\wedge\beta\rightarrow\gamma$ implies [redacted]$\wedge\beta\rightarrow\alpha\wedge\gamma$.

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 γ=β).

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.

---

terminology employed here better reflects the meaning of the rules.

Conditionalisation $$\frac{\alpha \mid\!< \beta\wedge\gamma}{\beta\rightarrow\alpha \mid\!< \gamma}$$

After having discussed various abstract properties of formation 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 formation 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 4.5. An *explanation mechanism* is some consequence relation $\mid\!\sim$. The *explanatory consequence relation* $\mid\!<$ *defined by* $\mid\!\sim$ is defined as $\alpha \mid\!< \beta$ iff $C_{\mid\sim}(\alpha) \subseteq C_{\mid\sim}($[redacted] L. A *strong* explanatory [redacted]onsequence relation is defined by a monotonic explana[redacted]n mechanism.

Thus, an explanation is required t[redacted]ave at least the same consequences under the explanation [redacted] as the premiss it is obtained from[redacted]nt. It should be noted that, in [redacted] general case, the conditions $C_{\mid\sim}(\alpha)\subseteq C_{\mid\sim}(\beta)$[redacted]nd $\beta \mid\!\sim \alpha$ are not equivalent.[15] However, for monotonic e[redacted]lanation mechanisms they are, which prov[redacted]es us with the following 'Peircean' definition of strong e[redacted] consequence relations.

DEFINITION 4.6. A *strong* [redacted]*planatory structure* is a set $W\subseteq M$. The consequence relati[redacted] it defines is denoted by $\mid\!<_W$ a[redacted] is defined by: $\alpha \mid\!<_W \beta$ iff (*i*) there is a $m_0\in W$ such that m_0 [redacted]β, and (*ii*) for every $m\in W$, m [redacted]$\beta\rightarrow\alpha$.

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

DEFINTION 4.7. The system **EM** consists of the following rules: Admissible Right Strengthening, Explanatory Ref[redacted]ivity, Incrementality, Predictive Convergence, Left Consistency, and Conditionalisati[redacted]

We note the following derived rules of [redacted]M: Convergence, Admissible Converse Entailment and[redacted] Reflexivity (instances of Predictive Con[redacted]emma 4.3). The following d[redacted] rule will also prove useful.

LEMMA 4.8. *The following rule is* [redacted]

Consistent Right Strengtheni[redacted] $$\frac{\alpha \mid\!< \gamma,\ \neg\beta \not\mid\!< \gamma}{\alpha \mid\!< \beta\wedge\gamma}$$

Proof. Suppose $\neg\beta \not\mid\!< \gamma$; since [redacted]$\neg(\beta\wedge\gamma)\wedge\gamma\rightarrow\neg\beta$, we have $\neg(\beta\wedge\gamma) \not\mid\!< \gamma$ by Predictive Convergence. Furthermore, sup[redacted]e $\alpha \mid\!< \gamma$, then by Right Reflexivity $\gamma \mid\!< \gamma$, so by Explanatory Reflexivity we have $\beta\wedge\gamma \mid\!< \beta\wedge\gamma$. We conclude by Admissible Right Strengthening.

Soundness of **EM** is easily checked.

LEMMA 4.9 (Soundness of **EM**). *Any strong explanatory consequence relation satisfies the rules of **EM**.*

[15] $C_{\mid\sim}(\alpha)\subseteq C_{\mid\sim}(\beta)$ implies $\beta \mid\!\sim \alpha$ if $\mid\!\sim$ is reflexive; $\beta \mid\!\sim \alpha$ implies $C_{\mid\sim}(\alpha)\subseteq C_{\mid\sim}(\beta)$ if $\mid\!\sim$ is transitive.

Proo[redacted]Let $W \subseteq U$ [redacted] stro[redacted]as
defi[redacted]d in Defini[redacted].6, sa[redacted]sfie[redacted] rules o[redacted]EM[redacted]
Adm[redacted]sible Right[redacted]engthening[redacted] W satisfies γ. Furthermore,
if m [redacted]β→α and [redacted]β, the[redacted]
Expl[redacted]atory Reflexi[redacted]y: w[redacted]ave t[redacted] some[redacted]ode[redacted] W sati[redacted]α, while not all models in
W satisfy α→¬β, i.e.[redacted]here [redacted]a mo[redacted] in W s[redacted]isfy[redacted] α∧β a[redacted]nce β.
Incrementality: if m [redacted]γ→α[redacted]nd m [redacted]γ→β th[redacted]n m [redacted]γ→(α∧[redacted].
Predictive Converge[redacted]ce: if [redacted]∧γ→β and m [redacted]γ→α then m [redacted]γ[redacted]β.
Left Consistency: if some [redacted]odel in W satisfi[redacted] β while all mo[redacted]
there is a model in W not satisfying β→¬α.
Conditionalisation: trivial.

In order to prove completeness we build a strong explanatory stru[redacted]ence relation $\mathrel{|\!<}$ satisfying the rules of **EM**, such that $\alpha \mathrel{|\!<} \beta$ iff $\alpha \mathrel{|\!<}_W \beta$. For n[redacted]-empty[redacted]xplanatory relations the following construction is used:

W = {$m \in U$ | for all α, β such that $\alpha \mathrel{|\!<} \beta$: [redacted]β→α[redacted]}

An empty explanatory relation signals inconsis[redacted]nt background k[redacted]wledge, [redacted]d is hence defined [redacted] empty explanatory structure.

We need a few intermediate results. The [redacted]llowing lemma [redacted]ates th[redacted] every strong expla[redacted] hypothesis is satisfiable in W.

LEMMA 4.10. *Let $\mathrel{|\!<}$ be a consequence [redacted]lation satisfying [redacted] defined as above. If $\alpha \mathrel{|\!<} \beta$ then there is a[redacted]odel $m \in W$ such [redacted]at m [redacted]β.*
Proof. Let $\alpha \mathrel{|\!<} \beta$; we will prove that {β}[redacted]ble. Suppose not, then by compactness there is a finite Δ⊆{δ→[redacted]$\gamma \mathrel{|\!<} \delta$} such th[redacted]→¬Δ. Furthermore, since $\varphi \mathrel{|\!<} \psi$ for any ψ→φ∈Δ, we have $\psi \to \varphi \mathrel{|\!<}$ [redacted]**ue** for any ψ→[redacted]Δ by Conditionalisation, $\Delta \mathrel{|\!<}$ **true** by Incrementality, and $\Delta \mathrel{|\!<} \beta$ [redacted]by Right Ref[redacted]xivity and Admissible Right Strengthening. But then by Consistency [redacted]β→¬Δ, a cont[redacted]

Furthermore, we have that every inadmissible formula is unsatis[redacted]able in W.

LEMMA 4.11. *Let $\mathrel{|\!<}$ be a non-empty consequence relati[redacted] **EM**, and let W be defined as above. If $\gamma \mathrel{|\!\not<} \gamma$ then γ is unsatisfiable [redacted]W.*
Proof. Let $\alpha \mathrel{|\!<} \beta$, then **true** $\mathrel{|\!<}$ **true** by Convergence and [redacted]ft Reflexivity. Furthermore, if $\gamma \mathrel{|\!\not<} \gamma$ then $\neg\gamma \mathrel{|\!<}$ **tr**[redacted] by Explan[redacted]ory Reflexivity, hence m [redacted]**rue** →¬γ for every $m \in W$.

I will now show that W[redacted]efines a co[redacted]equence relation that is included in $\mathrel{|\!<}$.

LEMMA[redacted].12. *L[redacted] satisfying the rules of **EM**, and let W b[redacted]efined[redacted]s abov[redacted] If [redacted]$<_W \beta$ then $\alpha \mathrel{|\!<} \beta$.*
Proof. [redacted]ppose [redacted]at α [redacted]β, [redacted]ill show that ei[redacted]er no model in W satisfies β, or there exists a[redacted]
First o[redacted]ll, if [redacted]β [redacted]unsatisfiable in[redacted]11. In the [redacted]maind[redacted] of th[redacted]roof [redacted]w[redacted]sume that $\beta \mathrel{|\!<} \beta$. [redacted]efine Γ_0 = {¬α} ∪ {δ | $\delta \mathrel{|\!<} \beta$}; we will [redacted]st sho[redacted] that [redacted]tness there is a finite Δ⊆{δ | δ [redacted]β} su[redacted] $\Delta \to \alpha \mathrel{|\!<} \beta$ [redacted]call t[redacted]t $\beta \mathrel{|\!<}$ [redacted]. But b[redacted] Incre[redacted]entality $\Delta \mathrel{|\!<} \beta$; us[redacted]g Incrementality and Convergence, we [redacted]t m_0 [redacted]Γ_0; c[redacted]rly m_0[redacted]α an[redacted] since β∈Γ_0, m_0 [redacted]β. It remains to prove that m_0 is in W; [redacted]., tha[redacted]or all[redacted], ψ su[redacted] that [redacted]ψ we have m_0 [redacted]ψ→φ. Let $\varphi \mathrel{|\!<} \psi$; if $\neg\beta \mathrel{|\!\not<} \psi$, then by [redacted]nsistent Rig[redacted] Strengthenin[redacted] and by Conditionalisation $\psi \to \varphi \mathrel{|\!<} \beta$; thus [redacted]→φ∈$\Gamma_0$ and[redacted]herefore m_0 [redacted]ψ → φ. On the other hand, if $\neg\beta \mathrel{|\!<} \psi$ then by [redacted]nditionalisat[redacted]n and Converg[redacted]ce $\beta \to \neg\psi \mathrel{|\!<}$ **true**, by Admissible Right Strengthening [redacted]→¬ψ $\mathrel{|\!<} \beta$, an[redacted]y Incrementality and Convergence $\neg\psi \mathrel{|\!<} \beta$; thus ¬ψ∈$\Gamma_0$ and therefore m_0 [redacted]ψ, hence m_0 [redacted]ψ→φ.

Armed with the previous three lemmas we can prove the co [redacted] pleteness [redacted] **EM**.

THEOREM 4.13 (Representation theorem for strong [redacted] xplanator [redacted] consequence relations). *A consequence relation is strong explanatory iff it sati* [redacted] *es the rul* [redacted]
Proof. The only-if part is Lemma 4.9. For the if [redacted] on-empty consequence relation satisfying the rules of **EM**, an [redacted] t

$$W = \{m\in U \mid \text{for all } \alpha, \beta \text{ such tha} \text{ [redacted] } \beta\text{: } m \text{ [redacted] } \beta\to\alpha\}$$

Suppose $\alpha \mathrel{|\!\!\!<} \beta$, then by the construction of W, m [redacted] $\to\alpha$ for all $m\in W$. Furthermore, by Lemma 4.10 there is a model in W satisfying β. We [redacted] y conclude that $\alpha \mathrel{|\!\!\!<}_W \beta$. Conversely, if $\alpha \mathrel{|\!\!\!<}_W \beta$ then Lemma 4.12 proves that $\alpha \mathrel{|\!\!\!<} \beta$. We conclude that W defines a consequence relation that is exactly $\mathrel{|\!\!\!<}$.
For an empty consequence relation put $W=\varnothing$.

In this section we have studied axioms and semantic characterisations for explanatory induction. I have proposed a novel definition of explanatory hypothesis formation 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 $\alpha \mathrel{|\!\!\!<} \beta$ is to be read as 'evidence α confirms hypothesis β'. Our goals will be to find reasonable properties of $\mathrel{|\!\!\!<}$ 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 [redacted]) into rules for confirmatory consequence relations. The conditions will be [redacted] ightly modified, in order to keep the treatment of inconsistent evidence and hypothesis in line with [redacted] e explanatory case: inconsistent evidence does not confirm any hypothesis, and inconsistent hyp [redacted] y any evidence.
Entailment condition (H1) is translated into two rules:

Admissible Entailment [redacted] $\dfrac{\to\beta\ ,\ \alpha \mathrel{|\!\!\!<} \alpha}{\alpha \mathrel{|\!\!\!<} \beta}$

Confirmatory Reflexivity $\dfrac{\alpha \mathrel{|\!\!\!<} \alpha\ ,\ \alpha \not\mathrel{|\!\!\!<} \neg\beta}{\beta \mathrel{|\!\!\!<} \beta}$

Admisible Entailment expresses that admissible evidence (i.e. evi [redacted] nce that is consistent with the background knowledge) confirms any of its consequences. In other [redacted] ords, consistent entailment is a special case of confirmation. Confirmatory Reflexivity is the confir [redacted] tory counterpart of Explanatory Reflexivity encountered in the previous section. It is added as a s [redacted] arate rule since, in its original formulation, (H1) includes reflexivity as a special case. As with its exp [redacted] Confirmatory Reflexivity is best understood when considering its contrapositive: if [redacted] ng a statement with regard to the background knowledge, its negation $\neg\beta$ [redacted] so weak that it is confirmed by arbitrary admissible formulae α.
Consequence condition (H2) cannot be translated directly, since [redacted] ence relations as defined here we have no means to refer to a set of [redacted] nfirmed sentences. However, a translation of the special consequence condition (H2.1) and the conjun [redacted] on condition (H2.3) will suffice:

Right Weakening [redacted] $\dfrac{\to\gamma\ ,\ \alpha \mathrel{|\!\!\!<} \beta}{\alpha \mathrel{|\!\!\!<} \gamma}$

Right And $$\frac{\alpha \mathrel{|\!\sim} \beta \,,\, \alpha \mathrel{|\!\sim} \gamma}{\alpha \mathrel{|\!\sim} \beta\wedge\gamma}$$

LEMMA 5.1. *If* $\mathrel{|\!\sim}$ *is a consequence relation satisfying Right A* [redacted] *and Right Weakening, then* $\alpha \mathrel{|\!\sim} \beta\wedge\gamma$ *iff* $\alpha \mathrel{|\!\sim} \beta$ *and* $\alpha \mathrel{|\!\sim} \gamma$,.
Proof. The *if* part is Right And, and the *only-if* part follows fr [redacted]

Right Weakening expresses that any hypothesis entailed by a gi [redacted] pothesis c [redacted] rmed by α [redacted] confirmed by α. Notice that Admissible Entailment is an instance o [redacted]

LEMMA 5.2. *The combination of Right Extension and Right* [redacted] *following rule:*

Predictive Right Weakening [redacted] $\beta\to\gamma \,,\, \alpha \mathrel{|\!\sim}$ [redacted]

Proof. In order to derive Predictive Right Weakening, supp [redacted] $\wedge\beta\to\gamma$ a [redacted] $\mathrel{|\!\sim} \beta$, then by Right Extension $\alpha \mathrel{|\!\sim} \beta\wedge\gamma$, and the result follows by Right [redacted] ening.
Predictive Right Weakening implies Right Weakening, since [redacted] $\to\gamma$ implies [redacted] $\beta\to\gamma$.
Predictive Right Weakening implies Right Extension, since [redacted] $\wedge\beta\to\gamma$ implies [redacted] $\wedge\beta\to\beta\wedge\gamma$.

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)[16]. It has already been remarked that Right And is probably [redacted] strong in the general case, if we have inconclusive evidence that is unable to choose between incomp [redacted] ble hypotheses. In this respect it is perhaps appropriate to point at a certain similarity between Right An [redacted] 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 co [redacted] ted directly into a rule, since we have no means to refer to the set of confi [redacted] ed formulae. However, in the light of Right And the conjunction of the formulae in this set is itsel [redacted] confirmed, and therefore it is sufficient to formulate a rule expressing the special consistency condit [redacted] of Consistency previously encountered:

Consistency [redacted] $\alpha \mathrel{|\!\sim} \beta$ / [redacted] $\beta\to\neg\alpha$

Condition (H3.2) expre [redacted] es that for a [redacted] formula β, if β is in the set of confirmed hypotheses then [redacted] not. This principle is ex [redacted]

Right Consi [redacted] **ency** $$\frac{\alpha \mathrel{|\!\sim} \beta}{\alpha \not\mathrel{|\!\sim} \neg\beta}$$

LEMMA 5.3. *I* [redacted] *the presen* [redacted] *of Admissible Entailment and Left Reflexivity, Right Consistency imp* [redacted] *s Consisten* [redacted].
Proof. Suppose [redacted] $\beta\to\neg\alpha$, i.e. [redacted] $\alpha\to\neg\beta$. Now, either we have $\alpha \mathrel{|\!\sim} \alpha$, or else $\alpha \not\mathrel{|\!\sim} \alpha$. In the former case, $\alpha \mathrel{|\!\sim}$ [redacted] β by Admis [redacted] ble Entailment, and we conclude by Right Consistency. In the latter case, we have $\alpha \not\mathrel{|\!\sim} \delta$ for any δ by Left Reflexivity.

Clearly, Consistency implies Right Consistency in the presence of Right And. As a corollary to Lemma 5.3, we have that Right Consistency and Consistency are equivalent in the presence of Left Reflexivity,

[16]This holds only for finite *K*, an assumption that I will make throughout.

Admissible Entailment, and Right And.

Finally, the equivalence condition for observations (H4) is translate[redacted]nto

Left Logical Equivalence $\dfrac{[redacted]\leftrightarrow\beta\,,\ \alpha \mathrel{|\!\!<} \gamma}{\beta \mathrel{|\!\!<} \gamma}$

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

5.2 *Simple confirmatory structures*

It has been suggested in section *2.5* that a sem[redacted]ics[redacted]co[redacted]irmatory reaso[redacted]g b[redacted]xpr[redacted]sed in terms of satisfaction by an appropriately constructed [redacted]de[redacted]se[redacted]of models. Mo[redacted]pr[redacted]ely[redacted]a confirmatory semantics is conceived as one in which certai[redacted]*egu*[redacted]mo[redacted]els are construc[redacted]d f[redacted]th[redacted]premisses, such that a hypothesis is confirmed if it is true [redacted]a[redacted]ost[redacted]regular models[redacted]Si[redacted]thi[redacted]requires some completeness assumptions regarding the evide[redacted]e [redacted]ay[redacted]all this a *close*[redacted]co[redacted]. In section *5.4* I will consider a variant which[redacted]la[redacted]the[redacted]assumptions re[redacted]rdi[redacted]th[redacted]ev[redacted]ce[redacted]*pen* confirmatory semantics).

DEFINITION 5.4[redacted]*c*[redacted]*ma*[redacted]*ry stru*[redacted]*re*[redacted]tr[redacted]e $W = \langle S,[\cdot], \cdot$[redacted]wh[redacted]$S$[redacted]a[redacted]of *semantic objects,*[redacted]nd[redacted]$L \rightarrow$[redacted]S and $\cdot$[redacted]$\rightarrow$[redacted]re[redacted]nctions mappi[redacted]fo[redacted]la[redacted]of semantic objects.[redacted]he[redacted]*ed*[redacted]nfirmat[redacted]c[redacted]qu[redacted]ce relation defi[redacted]d b[redacted]V[redacted]iv[redacted]y: $\alpha \mathrel{|\!\!<}_W \beta$ iff (*i*) $[\alpha]$[redacted]$\emptyset$,[redacted]β

Intuitively, $[\alpha]$ denote[redacted]the[redacted]of[redacted]ula[redacted]m[redacted]ted from premisses α, eac[redacted]of[redacted]h[redacted]ul[redacted] satisfy hypothesi[redacted]n c[redacted]dere[redacted] by Bell [2], who calls it [redacted]rag[redacted]ic m[redacted]del. There are howe[redacted]ach and mine. First, in[redacted]de[redacted]ru[redacted]out inconsistent pre[redacted]ion (*i*) in the definition of $\mathrel{|\!\!<}_W$. Further[redacted]re[redacted]llo[redacted]he possibility that s[redacted]e[redacted]ay not satisfy the premisses (i.e. $[\alpha]$ ¤ [redacted]o this possibility [redacted]fu[redacted]on **6**. However, the characterisation of[redacted]ive logics of confir[redacted]ati[redacted]roblem, and the results obtained belo[redacted]r[redacted]the assumption that [redacted]]⊆[redacted]results differ from those reported in [1[redacted]]s[redacted]the additional condi[redacted]n ([redacted]he present section and the next.

Bell proves [redacted]T[redacted]ragmatic models, including the con[redacted]io[redacted]]⊆[redacted]are axiomatised by [redacted]e r[redacted]g of Reflexivity, Right Weakening, an[redacted]Ri[redacted]An[redacted]one additionally ass[redacted]nes[redacted]ved with respect to the logical conn[redacted]tiv[redacted]nd[redacted]ca entailment:

α[redacted]=[redacted]
¬[redacted]=[redacted]
α[redacted]*S*[redacted]

Let us call confi[redacted]sfy these conditions, as well as $[\alpha] \subseteq \alpha$[redacted]or[redacted]$\in L$[redacted]*imple* confirmatory st[redacted]e (closed) confirmatory conseque[redacted]w[redacted]now demonstrate that[redacted]matory structures are axiomatised by the follow[redacted]g r[redacted]yst[redacted]

DEFINITION 5.5. The system **CS** consists of the following rule[redacted]Pr[redacted] Weakening, Right And, and Right Consistency.

Derived rules of **CS** include Right Weakening and Right Extension (Lemm[redacted]5.2[redacted]ft[redacted]flexivity (an instance of Predictive Right Weakening), Admissible Entailment (an instanc[redacted]f R[redacted]W[redacted]kening), and Consistency (Lemma 5.3).

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

LEMMA 5.6 (Soundness of **CS**). *Any simple closed confirmatory c*[redacted]*se*[redacted]*ce*[redacted]*lation satisfies the rules of* ***CS***.

Proof. For Predictive Right Weakening, first of all we have $[\alpha] \subseteq \alpha$[redacted]or[redacted]$\in$[redacted]hence [α]∩ β [redacted]

[α]∩ β [redacted]

and we [redacted]

For Rig[redacted]

For Rig[redacted]

In order to prove completeness w[redacted]e c[redacted]firmatory structure *W* that defines exactly th[redacted]

DEFINITION 5.7. Let $\mathrel{|\!<}$ be a confirmatory conseque[redacted]e relation. The model *m*∈ *U* is said to be *normal for* α iff for all β in *L* such that α $\mathrel{|\!<}$ β, *m* [redacted] β.

For ad[redacted]issible fo[redacted]ulae normal models will play the role of the regular models in the s[redacted] confirm[redacted]tory struct[redacted]e we are building (remember that, given a consequence relation $\mathrel{|\!<}$, a formula α [redacted] admiss[redacted]e iff α $\mathrel{|\!<}$ [redacted]

[redacted]EMMA [redacted] *consequence relation* $\mathrel{|\!<}$ *satisfies Right Weakening and Right And,* [redacted]*d let* α [redacted]*ble formula. All normal models for* α *satisfy* β *iff* α $\mathrel{|\!<}$ β.
[redacted] from Definition 5.7.
[redacted]pose α $\mathrel{|\!<}$ α and α $\not\mathrel{|\!<}$ β; I will show that there is a normal model for [redacted] that do[redacted]. Let Γ_0 = {¬β} ∪ {δ | α $\mathrel{|\!<}$ δ}; it suffices to show that Γ_0 is [redacted]tisfiabl[redacted] then by compactness there is a finite Δ⊆{δ | α $\mathrel{|\!<}$ δ} such that [redacted]Δ→β, i[redacted] by Right Weakening α $\mathrel{|\!<}$ Δ→β. But by Right And α $\mathrel{|\!<}$ Δ; using [redacted]kening we obtain α $\mathrel{|\!<}$ β, a contradiction.

In the standar[redacted]nconsistent premisses they have all formulae as consequences, hence no normal models [redacted]eatment inconsistent premisses confirm no hypothesis and thus have all models in *U* as [redacted]ve have to treat them as a separate case. Given a consequence relation $\mathrel{|\!<}$, let *W* = ⟨*U*,[·], · [redacted]
(1) *U* [redacted]els of *L* under consideration;
(2) [α[redacted]a normal model for α} if α is admissible, an[redacted]
(3) [redacted]
The following c[redacted]sult dem[redacted]he consequence rela[redacted]*W* coi[redacted] with the original[redacted]r satisfies[redacted].

THEORE[redacted]entation th[redacted]ple confirmatory cons[redacted]ons). *A consequence relation is simple co*[redacted] *satisfies the rules of* [redacted].
Proof. The only-if part is Lemm[redacted]e if part, let $\mathrel{|\!<}$ be a [redacted]elation satisfying the rules of **CS** and let [redacted]s above. We will pro[redacted] α $\mathrel{|\!<}_W$ β, i.e. $\mathrel{|\!<}$ is simple confirmatory.
First suppose that α $\mathrel{|\!<}$ β, then by [redacted] α $\mathrel{|\!<}$ α an[redacted] $\not\mathrel{|\!<}$ ¬α, so the proof of Lemma 5.8 constr[redacted] normal fo[redacted]ermore any normal model for α satisfies [redacted] 5.7, henc[redacted] that α $\mathrel{|\!<}_W$ β. Now suppose that α $\mathrel{|\!<}_W$ β, t[redacted]nce α is ad[redacted] of *W*. Furthermore we have [α] ⊆ β[redacted]nd the conclusion follows by Lemma 5.8[redacted]

One may note that two of the rules ob[redacted]previous se[redacted]ed above, *viz.* Confirmatory Reflexivity and Lef[redacted]ivalence. [redacted]ules poses additional restrictions on simple confirmatory structures:

(Left Logical Equivalence) If [redacted]β, [redacted] = [β].

(Confirmatory Reflexivity) If β[redacted]≠ ∅.

Additional properties may be obtained by being more explicit abou[redacted]ion of regular models. The ILP methods referred to in section *2.5* suggest to take the trut[redacted]brand model(s) of the evidence as the regular model(s). In the analysis of nonmonotonic [redacted] 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 [redacted]

5.3 Preferential [redacted]

The main result [redacted] of Kraus [redacted] semantics, [redacted] confirmatory [redacted] semantic [redacted] that are [redacted] respect to a fixed [redacted]

> [redacted] is [redacted] where S is a set of [redacted] l: $S \rightarrow U$ is [redacted] a model, [redacted] partial order[17] [redacted] S, called th[redacted]. W de[redacted] *structure* ⟨[redacted] ry conse[redacted] n as follows: α [redacted] $\{s \in S \mid l(s)$ [redacted] $\alpha\}$, a[redacted] $s' \notin$ α [redacted]

Note that preferential c[redacted] ple con[redacted] tures, and that [redacted] satisfy the conditions ass[redacted] Logical Equiv[redacted] Confirmatory Refle[redacted] the smoothness condition, if $s \in$ α [redacted] r there is [redacted] $t \in [\alpha]$ — hence α [redacted] $\neq \emptyset$ implies $[\alpha] \neq \emptyset$).

In comparison with the prefere[redacted] s of [18], the only difference is that in a prefer[redacted] confirmatory argument the evide[redacted] to be satisfiable, in order to guarantee the valid[redacted] Consistency. The intermediate sem[redacted] *states* is mainly needed for technical reasons, and c[redacted] interpreted as 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 5.11. The system **CP** consists of the rules of **CS**, Confirmatory Reflexivity, Left Logical Equivalence, plus the following rules:
>
> **Left Or** $$\frac{\alpha \mathrel{|\!\!<} \gamma ,\ \beta \mathrel{|\!\!<} \gamma}{\alpha\vee\beta \mathrel{|\!\!<} \gamma}$$
>
> **Strong Verification** $$\frac{\alpha \mathrel{|\!\!<} \gamma ,\ \alpha \mathrel{|\!\!<} \beta}{\alpha\wedge\gamma \mathrel{|\!\!<} \beta}$$

The first [redacted] eneral case: if we weaken t[redacted] r confirms the hypothesi[redacted] by taking [redacted] formula γ [redacted] also allow[redacted] very simi[redacted] evidence [redacted]

[redacted]

---

[17]I.e., < is irreflexive a[redacted]

[18]I.e. for any $S' \subseteq S$ and for any $s \in S'$, either s is minimal in S', or there i[redacted] S'. This condition is satisfied if < does not allow infinite descending chains.

[19]In 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.

non-empty, hence $[\alpha\wedge\gamma] \neq \emptyset$. N [redacted] then $s\in$ α [redacted]
Suppose not, then there is a $t\in$ α [redacted] $t\in[\alpha]\subseteq$ [redacted]
contradicts the minimality of s in [redacted]

In order to prove completeness, we ne [redacted] ential stru [redacted] relation $\mid\!\sim$ satisfying the rules of **CP**, such [redacted] as follows:

(1) $S = \{\langle m,\alpha\rangle \mid \alpha$ is an admissible fo [redacted] ula, and m is a normal model for $\alpha\}$;
(2) $l(\langle m,\alpha\rangle) = m$;
(3) $\langle m,\alpha\rangle < \langle n,\beta\rangle$ iff $\alpha\vee\beta \mid\!\sim \alpha$ and m [redacted] β.

Thus, states are pairs of admissible formulae a [redacted] normal models. The labelli [redacted] simply [redacted] state to the model it contains. Condition (3) defines [redacted] ordering [redacted] twe [redacted] sta [redacted]: note t [redacted] α is a special case of $\alpha\vee\beta \mid\!\sim \alpha$ by means of Left Or, [redacted] nd [redacted] act [redacted] at α is ad [redacted] ssi [redacted] Th [redacted] ondition [redacted] is added to make the ordering irreflexive; note that [redacted] a [redacted] equ [redacted] ce any ⟨ [redacted] α⟩∈ [redacted] mi [redacted] al in α [redacted]

The main difference between the preferential co [redacted] eq [redacted] r [redacted] tions of [redacted] *l.* a [redacted] my prefe [redacted] ti [redacted] confirmatory consequence relations is the way u [redacted] tisf [redacted] fo [redacted] nulae are [redacted] eat [redacted] K [redacted] us *et al.*'s [redacted] te [redacted] **P** unsatisfiable formulae are characterised by th [redacted] fa [redacted] t t [redacted] y have ev [redacted] y [redacted] ula [redacted] L as a pl [redacted] sib [redacted] consequence, which means that they don't have [redacted] od [redacted]. In my f [redacted] ne [redacted], u [redacted] atisfiable fo [redacted] confirm no hypotheses, and have all models in [redacted] al [redacted] al [redacted] odels. In [redacted] oth [redacted] s, t [redacted] structure W [redacted] tha [redacted] used to prove completeness contains only sa [redacted] fi [redacted] for [redacted] ulae in it [redacted] sta [redacted] Th [redacted] means tha [redacted] ve [redacted] replicate most of Kraus *et al.*'s results about th [redacted] st [redacted]

PROPOSITION 5.13. *(1) [18, Lemma 5* [redacted] *]* [redacted] *rel* [redacted] *on < is a* [redacted] *ic* [redacted] *tia* [redacted] *rder.*
(2) [18, Lemma 5.15] The relation < [redacted] *s* [redacted] *fo* [redacted] *ny s∈* α [redacted] *ith* [redacted] *is* [redacted] *imal in* α [redacted] *r there exists a state t<s minimal in* α [redacted]
(3) [18, Lemma 5.11] If $\alpha\vee\beta \mid\!\sim \alpha$ *an* [redacted] *orr* [redacted] *l model fo* [redacted] *s β, then* [redacted] *s a* [redacted] *no* [redacted] *r β.*
(4) [redacted] *18,* [redacted] *m* [redacted] *.14]* $\langle m,\alpha\rangle\in[\beta]$ *i* [redacted] *nd* [redacted] $\beta \mid\!\sim \alpha$.

The first [redacted] o [redacted] me [redacted] s express that W is a preferential structure. [redacted] he [redacted] ain [redacted] g two are used [redacted] proof of t [redacted] fo [redacted] ing [redacted] mma.

LE [redacted] M [redacted] 14. [redacted] *et* $\mid\!\sim$ *be a consequence relation satisfying th* [redacted] *rul* [redacted] ***C*** [redacted] *and let W be de* [redacted] *ed* [redacted] *bo* [redacted] *If* $\alpha \mid\!\sim \beta$ *then* $\alpha \mid\!\sim_W \beta$.
Pr [redacted] f. [redacted] sup [redacted] ose that $\alpha \mid\!\sim \beta$, then by Le [redacted] Ref [redacted] ivit [redacted] $\mid\!\sim$ α [redacted] nd [redacted] igh [redacted] Consistency α $\mid\!\sim$ [redacted] α, [redacted] he [redacted] roof of Lemma 5.8 const [redacted] cts [redacted] mod [redacted] m n [redacted] rm [redacted] r [redacted] hence $[\alpha]\neq\emptyset$. Fu [redacted] her [redacted] e su [redacted] pose $s=\langle n,\gamma\rangle\in[\alpha]$, then γ is [redacted] n a [redacted] ssib [redacted] form [redacted] la, [redacted] a n [redacted] rma [redacted] γ t [redacted] t sa [redacted] es [redacted] and $\gamma\vee\alpha \mid\!\sim \gamma$ [redacted] Propositi [redacted] 5.1 [redacted]. B [redacted] Propositi [redacted] 13 [redacted]) n [redacted] a n [redacted] al mo [redacted] el f [redacted] a [redacted] thus satisfies [redacted] by Definiti [redacted] 5. [redacted] nc [redacted] s∈ β [redacted] We [redacted] clu [redacted] tha [redacted] ∅⊂ [redacted] ⊆ β [redacted] en [redacted] $\mid\!\sim$ [redacted] β.

The follo [redacted] ng [redacted] ma [redacted] roves the conv [redacted] mpletes [redacted] pr [redacted] f of [redacted] e r [redacted] se [redacted] theorem.

L [redacted] M [redacted] 5.15. *Let* $\mid\!\sim$ *be a con* [redacted] *quence re* [redacted] *ion* [redacted] *isf* [redacted] *g the rules of* ***CP****, an* [redacted] *let* [redacted] *e defined as above. If* $\alpha \mid\!\sim_W \beta$ *th* [redacted] $\alpha \mid\!\sim \beta$.
Proof. Suppose $\alpha \mid\!\sim_W \beta$, i.e. [redacted] $[\alpha] \subseteq$ β [redacted] we [redacted] fi [redacted] prove that α is admi [redacted] ibl [redacted] $\langle n,\gamma\rangle\in[\alpha]$, then $\gamma \mid\!\sim \gamma$ and n i [redacted] a normal [redacted] od [redacted] r [redacted] If α $\mid\!\sim$ α, then by Co [redacted] irm [redacted] y Reflexivity $\gamma \mid\!\sim \neg\alpha$ and thus n [redacted] ¬α, contra [redacted] cti [redacted] e [redacted] sumption that $\langle n,\gamma\rangle\in[\alpha]$ [redacted] s admissible. Furthermore, give [redacted] any mod [redacted] m [redacted] al [redacted] r α, $\langle m,\alpha\rangle\in[\alpha] \subseteq$ β [redacted] he [redacted] m satisfies β, and the conclusion follows by L [redacted]

We may now summarise.

THEOREM 5.16 (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 5.12. For the if part, let $\mid\!\sim$ be a consequence relation

satisfying the rules of **CP** and let W be defined as above. Lemma 5.14 and Lemma 5.15 prove that $\alpha \mathrel{|\!\!<} \beta$ iff $\alpha \mathrel{|\!\!<}_W \beta$, i.e. $\mathrel{|\!\!<}$ 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 $\mathrel{|\!\!\sim}$ we can construct a preferential confirmatory consequence relation by removing all arguments $\{\alpha \mathrel{|\!\!\sim} \beta \mid \alpha \mathrel{|\!\!\sim} \delta$ for all $\delta \in L\}$, i.e. all arguments with a left-hand side that is inconsistent with respect to the background knowledge. The resulting preferential confirmatory relation is the largest one contained in $\mathrel{|\!\!\sim}$; furthermore, all preferential confirmatory relations can be constructed in this way from a preferential relation. Thus, **CP** is more restrictive than **P**[redacted]efined at the end of section *3.3*.

5.4 *Weak confirmatory consequence relati*[redacted]*s*

Any se[redacted]an[redacted]th[redacted] is to obey rules lik[redacted]Ri[redacted]A[redacted] and Strong V[redacted]ifi[redacted]n[redacted]ust be based on comple[redacted]nes[redacted]su[redacted]ptions with regard to [redacted]e e[redacted]en[redacted] On the other [redacted]nd[redacted]ch[redacted]ong assumptions cannot [redacted]e n[redacted]uction tasks. It [redacted]een[redacted]rea[redacted]nable, then, to[redacted]ve[redacted]te[redacted]so an alternative approa[redacted]in[redacted]ned hypothesis i[redacted]eq[redacted]d t[redacted]e true in *some* [redacted]th[redacted]ula[redacted]models.[20] It is not difficul[redacted]o s[redacted]lternative sema[redacted]ics[redacted]sed[redacted]n some notion [redacted]co[redacted]en[redacted] would invalidate both Ri[redacted]t A[redacted]Verification. On[redacted]e o[redacted] ha[redacted]d, by not makin[redacted]co[redacted]en[redacted]s assumptions one may ag[redacted]h[redacted]e property of Co[redacted]erg[redacted]e.

D[redacted]FI[redacted]t $W = \langle S,[\cdot], \cdot$ [redacted]e a[redacted]nfi[redacted]atory structure[redacted]he[redacted]*n*[redacted]nfirmatory [redacted]nse[redacted]defined by W is[redacted]ve[redacted]: [redacted]$_W$ β iff $[\alpha]\cap \beta$[redacted]

We wil[redacted]ha[redacted]eme form of op[redacted]c[redacted]rma[redacted]ry relations, w[redacted]ch[redacted]s w[redacted]en $[\cdot]$ is identified with $\cdot$ [redacted]hi[redacted]-behaved with [redacted]connectives and[redacted]nt.

D[redacted]FI[redacted] *classical confirmatory structure* is a simple confirmatory structure ⟨[redacted]ce relation is called weak confirmatory iff it is the open consequence [redacted]lassical confirmatory structure.

From this [redacted]clear that weak confirmatory consequence relations satisfy both Right Weakening [redacted]ening (i.e. Convergence), as well as Consistency. One additional rule is needed.

DEFINITION 5.19. The system **CW** consists of the following rules: Predictive Convergence, Predictive Right Weakening, Consistency, and

Disjunctive Rationality $\dfrac{\alpha\vee\beta \mathrel{|\!\!<} \gamma\,,\ \beta \not\mathrel{|\!\!<} \gamma}{\alpha \mathrel{|\!\!<} \gamma}$

Disjunctive Rationality has not been considered before. T[redacted]n borrowed from Kraus *et al.*, who identify it as a valid principle of plausible reasoni[redacted]xt of confirmatory reasoning, Disjunctive Rationality is a rather strong rule, which [redacted] hypothesis is confirmed by disjunctive observations it is confirmed by at least one of t[redacted]

The following theorem proves the equivalence of weak [redacted]ctures and the system **CW**.

THEOREM 5.20 (Representation theorem for weak [redacted]nsequence relations). *A consequence relation $\mathrel{|\!\!<}$ is weak confirmatory iff it s*[redacted]*of* ***CW***.

Proof. The only-if part involves demonstrating th[redacted]ed in Definition 5.18, satisfies the rules of **CW**, which is trivial.

For the if part, let $\mathrel{|\!\!<}$ be an arbitrary consequence re[redacted] the rules of **CW**, and consider the weak confirmatory structure $W = \langle S, \cdot$ [redacted]

[20]This is sometimes called *credulous* inference, in contrast with [redacted] which requires truth in all regular models.

[redacted] S[redacted]m[redacted]for[redacted],$\beta\in L$ such that m[redacted]$\alpha\wedge\beta$: $\alpha \mathrel{|\!<} \beta$[redacted]

[redacted] {r[redacted]| m[redacted]

> We will[redacted]ve[redacted]α[redacted]β[redacted]f $\alpha \mathrel{|\!<}_W \beta$. The if pa[redacted] follows [redacted]ectly [redacted]
> W. Supp[redacted] mo[redacted]l $m_0\in S$[redacted]
> Define [redacted]$\not\mathrel{|\!<} \beta$}; we will first [redacted]
> by com[redacted]nite $\Delta\subseteq${δ | $\neg\delta\not\mathrel{|\!<} \beta$[redacted]such that [redacted]$\neg(\Delta\wedge\alpha)$[redacted].e. [redacted]$(\Delta\wedge\alpha)$; by
> Consist[redacted], we have[redacted]
> Rationa[redacted]
> $(\neg\Delta\wedge\alpha)$[redacted]$(\Delta\wedge\alpha) \not\mathrel{|\!<} \beta$ by D[redacted]unctive Rationality[redacted]ntradiction,
> so Γ_0 is[redacted]tisfiable.
> Let m_0[redacted]Γ_0; clearly m_0[redacted]α and, since by C[redacted]istency $\beta\in\Gamma_0$, m_0[redacted]β. It rem[redacted]ins to prove
> that m_0[redacted]; i.e., that for a[redacted]φ, ψ such that m_0[redacted]$\varphi\wedge\psi$ we have $\varphi\not\mathrel{|\!<}$[redacted]. Let m_0[redacted]$\varphi\wedge\psi$, then
> $\neg(\varphi\wedge\psi)\notin\Gamma_0$, hence $\varphi\wedge\psi \mathrel{|\!<} \beta$; by Predicti[redacted]e Right Weakening $\varphi\wedge\psi$[redacted]$\psi\wedge\beta$, by
> Convergence $\varphi \mathrel{|\!<} \psi\wedge\beta$, and by Right Weakening $\varphi \mathrel{|\!<} \psi$.

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 more meaningful forms of open confirmatory reasoning.

6. Discussion

In this paper 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 paper is written there is not one, but two 'problems of induction': one concerned with the justification of accepted hypotheses, and one concerned with the formation of possible 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.' [21, Book III, Chapter VIII, p.314]

No attempt has been made, in the present paper, 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 paper, is the problem of finding a sufficiently accurate description, in logical terms, of the process of inductive hypothesis formation. 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”[21], 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.’ [12, pp.1073–4]

The first philosophical contribution of this paper, then, has been to reinforce the point that inductive hypothesis formation is a logical matter (and hypothesis 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 formation 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 set up, 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 paper to predicate logic is certainly an important open problem. On the other hand, the propositional analysis of this paper is not meaningless for the predicate logic case. First of all, in finite domains statements of predicate logic can be encoded in propositional logic.[22] 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 [redacted]s the (restricted) reflexivity of the logics proposed here, which again runs counter to th[redacted]nduction as generalisation from instances to populations. We expect names of individuals [redacted]in the inductive premisses, and absent from the inductive hypothesis. As a consequenc[redacted] names in the premisses (in a way that does not distort the information they convey) wo[redacted]te the hypothesis. For instance, consider the observations

[redacted]ack(a) ∧ ¬Raven(c)∧Black(c) ∧ ¬Raven(d)∧¬Black(d)

Given these[redacted]∀x: Raven(x)→Black(x) is a possible hypothesis. In the confirmatory setting this can be[redacted] constructing a regular model from the evidence, and stipulating that any hypothesis [redacted] that model. However, from that model we can construct another model by replacing a,[redacted] e, f, g. It makes sense to say that this constitutes another regular model that is to satisfy a[redacted]— this would rule out any hypothesis that talks about a, c or d, including the observations[redacted]e confirmatory structures of section *5.2* this would mean to drop the condition that [α]⊆ α[redacted]n this way the perception of induction as reasoning from the particular to the general wo[redacted] not by posing syntactic restrictions (which is logically unattractive), but by wiring it int[redacted]. I am currently working on such a logic of generalisation.

6.2 Logic

Mathematical logic has been at the focus of attention of logicians at least since the beginning of the century. This has led to an underappreciation of other forms of reasoning. Work on commonsense

[21] Peirce’s translation of Aristotle’s term απαγωγη — only later Peirce introduced the term ‘abduction’.
[22] This may require a substantial background theory, limiting the practical feasibility of this encoding.

reasoning in Artificial Intelligence has revived the interest in non-standard logics, but still most of the map of reasoning forms is in darkness. The regrettable use of the term ‘nonmonotonic reasoning’ for reasoning with rules that have exceptions[23] is a symptom of this twilight: virtually all nondeductive reasoning is nonmonotonic, yet reasoning with default rules constitutes but one possible mode of nondeductive reasoning. Logic should also be concerned with the systematic study of *reasoning forms* (deduction, induction, plausible reasoning, counterfactual reasoning, and so on).

From a logical point of view this paper 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 analyis, 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 [18] chart the map of plausible reasoning. Thus, the present paper can be seen as a constructive proof of the thesis that the method of analysis through consequence relations, pioneered by Gabbay [11], Makinson [20], and Kraus, Lehmann & Magidor [18, 19], 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.’ [17, 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’ [25, 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 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 explanatory consequence relations studied in section **4** of this paper satisy none of these properties. This paper makes a case for a much more liberal perception of what is a consequence relation.

---

[23] A better term would be ‘plausible reasoning’.

6.3 Machine Learning

Whereas the main contributions of this paper 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) [22]. 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 paper. 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 *2.5*), 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 insig[redacted] in and justification of the new concept. For instance, the definition of a definite integral as the limit[redacted] a Riemann sum formalises the idea that a definite integral calculates the area under a curve. The [redacted]lationship between the new concept and previously defined concepts is then obtained as a theorem [redacted]. the fundamental theorem of calculus), rather than a seemingly arbitrary definition.

The framework of inductive consequence rela[redacted]ns can be used to obtain soundness and completeness proofs of particular sets of operators employed [redacted] a particular induction algorithm. As an illustration, consider the inverse resolution operator absorptio[redacted]–`q,r,t` infer `s←p,t`. In our framework, a soundness proof of this ind[redacted]tion step is established by the derivation, from an appropriate rule system such as **EM**, of the follo[redacted]g statement:

$$\frac{[redacted]\leftarrow \mathtt{q,r}}{\mathtt{s}\leftarrow\mathtt{q,}[redacted]\mathtt{,t} \mid\!\prec \mathtt{s}\leftarrow\mathtt{p,t}}$$

which should be read as follows: if `p←q,r` is known from the background knowledge, then `s←p,t` is an inductive consequence of `s←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 paper: 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 [14, 6, 5]. The contribution of the present paper 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 paper 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 formation.

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 paper I have proposed a metalevel framework for characterising and reasoning about different forms of induction. The framework does not fix a material definition of inductive hypothesis formation, 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 [2] and Kraus, Lehmann & Magidor [18], 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 PECO Pan-European Scientific Network ILPnet (no. CIPA3510OCT920044).

References

[1] D. Angluin & C.H. Smith (1983), ‘Inductive inference: theory and methods’, *Computing Surveys* **15** (3): 238–269.
[2] J. Bell (1991), ‘Pragmatic logics’. In *Proc. Second International Conference on Knowledge Representation and Reasoning KR’91*, Morgan Kaufmann, San Mateo, pp. 50–60.
[3] R. Carnap (1950), *Logical Foundations of Probability*, Routledge & Kegan Paul, London.
[4] J.P. Delgrande (1987), ‘A formal approach to learning from examples’, *Int. J. Man-Machine Studies* **26**, pp. 123-141.
[5] L. De Raedt & M. Bruynooghe (1993), ‘A theory of clausal discovery’. In *Proc. 13th International Conference on Artificial Intelligence IJCAI’93*, Morgan Kaufmann, San Mateo, pp. 1058–1063.
[6] P.A. Flach (1990), ‘Inductive characterisation of database relations’. In *Proc. International Symposium on Methodologies for Intelligent Systems ISMIS’90*, Z.W. Ras, M. Zemankowa & M.L. Emrich (eds.), pp. 371-378, North-Holland, Amsterdam. Full version appeared as ITK Research Report no. 23.
[7] P.A. Flach (1993), ‘Predicate invention in Inductive Data Engineering’. In *Proc. European Conference on Machine Learning ECML’93*, P.B. Brazdil (ed.), Lecture Notes in Artificial Intelligence 667, Springer-Verlag, Berlin, pp. 83–94.

[8] P.A. Flach (1995), *Conjectures: an inquiry concerning the logic of induction,* PhD thesis, Tilburg University.

[9] P.A. Flach (1996), 'Comparing consequence relations', manuscript.

[10] P.A. Flach (forthcoming), 'Logical approaches to abductive reasoning and learning: an overview', forthcoming.

[11] D.M. Gabbay (1985), 'Theoretical foundations for non-monotonic reasoning in expert systems'. In *Logics and Models of Concurrent Systems,* K.R. Apt (ed.), Springer Verlag, Berlin: 439–457.

[12] N.R. Hanson (1958), 'The logic of discovery', *Journal of Philosophy* **55**(25): 1073–1089.

[13] C. Harstshorne, P. Weiss & A. Burks, eds (1931-58), *Collected Papers of Charles Sanders Peirce*, Harvard University, Cambridge.

[14] N. Helft (1989), 'Induction as nonmonotonic inference'. In *Proc. First International Conference on Knowledge Representation and Reasoning KR'89,* Morgan Kaufmann, San Mateo, pp. 149–156.

[15] C.G. Hempel (1943), 'A purely syntactical definition of confirmation', *Journal of Symbolic Logic* **6**(4): 122–143.

[16] C.G. Hempel (1945), 'Studies in the logic of confirmation', *Mind* **54**(213): 1–26 (Part I); **54**(214): 97–121 (Part II).

[17] R.E. Jennings, C.W. Chan & M.J. Dowad (1991), 'Generalised inference and inferential modelling'. In *Proc. Int. Joint Conf. on Artificial Intelligence IJCAI'91*, Morgan Kaufmann: 1046–1051.

[18] S. Kraus, D. Lehmann & M. Magidor (1990), 'Nonmonotonic reasoning, preferential models and cumulative logics', *Artificial Intelligence* **44**, pp. 167-207.

[19] D. Lehmann & M. Magidor (1992), 'What does a conditional knowledge base entail?', *Artificial Intelligence* **55**, pp. 1-60.

[20] D. Makinson (1989), 'General theory of cumulative inference', in *Proc. 2nd International Workshop on Non-Monotonic Reasoning*, M. Reinfrank, J. de Kleer, M.L. Ginsberg & E. Sandewall (eds.), Lecture Notes in Artificial Intelligence 346, Springer-Verlag, Berlin: 1–18.

[21] J.S. Mill (1843), *A System of Logic.* Reprinted in *The Collected Works of John Stuart Mill*, J.M. Robson (ed.), Routledge & Kegan Paul, London.

[22] S. Muggleton & L. De Raedt (1994), 'Inductive Logic Programming: theory and methods', *Journal of Logic Programming*.

[23] S. Muggleton & W. Buntine (1988), 'Machine invention of first-order predicates by inverting resolution', in *Proc. Fifth International Conference on Machine Learning*, J. Laird (ed.), Morgan Kaufmann, San Mateo: 339–352.

[24] A. Tarski (1936), 'Über den Begriff der logischen Folgering', *Actes du Congrès Int. de Philosophie Scientifique* 7: 1–11. Translated into English as 'On the concept of logical consequence'. In *Logic, Semantics, Metamathematics,* A. Tarski (1956), Clarendon Press, Oxford: 409–420.

[25] W. Zadrozny (1991), *On rules of abduction*, IBM Research Report, August 1991.