42

Modern Logic and its Role in the Study of Knowledge

PETER A. FLACH

Knowledge is at the heart of intelligent behavior. The ability to obtain, manipulate, and communicate knowledge, in explicit form, is what distinguishes humans from other animals. This suggests that any study of intelligent behavior, theoretical or experimental, would have the same starting point, namely a Science of Knowledge, which studies the basic forms of knowledge, its acquisition, and its processing.

Yet there does not seem to exist such a unified and mutually agreed science of knowledge. In ancient times philosophy, the 'love of knowledge,' would aim to fulfill this role of the Mother of all Sciences, but philosophy has since long lost its central place and has mostly fragmented into specialized sciences such as physics, biology, and mathematics. Computer science, a relatively young branch on the tree of knowledge, has some aspirations to be the science of knowledge, but is currently at best a loosely connected collection of engineering technologies and abstract mathematical theory. (In fact, scholars of more established disciplines such as physics or chemistry often hesitate to call computer science a science at all, because its design-oriented approach does not fit in well with the doctrines of experimental sciences.) Artificial intelligence – the discipline studying fruitful connections between intelligent behavior and computers – would be another contender, but has been accused of overstating its claims, having unclear goals, and applying sloppy methodology.

In this chapter I argue that logic, in its widest sense, is – or at least, should be perceived as – the science of knowledge. This would be an unsurprising statement for a nineteenth-century logician, who would study the kind of inductive reasoning involved in experimental sciences as eagerly as he would investigate the kind of reasoning that is employed in mathematical proofs. However, in the last century logic seems to have developed into a relatively specialized and not seldomly obscure branch of mathematics. This is all the more paradoxical since the first half of the twentieth century has often been called 'the Golden Age of logic.' Following the pioneering work of Gottlob Frege, who developed a forerunner of predicate logic called *Begriffsschrift* ('concept language') in 1893, Russell and Whitehead published their three-volume *Principia Mathematica* between 1910 and 1913, in which they re-established the foundations of pure mathematics in logical terms. Whereas Kurt Gödel dealt a severe blow to the ambitions of logicians when he demonstrated that any logical system powerful enough to include natural numbers is also necessarily incomplete (i.e. the logical system allows the formulation of true statements which are demonstrably unprovable within the system), this didn't stop logicians like Alonzo Church to develop ever more powerful logical systems (e.g. combinator logic and higher-order logic). Furthermore, Alfred Tarski invented what I consider one of the most important contributions of modern logic, namely the notion of an independent semantics.

1 The Key Ingredients of Logic

The duality between syntax and semantics is not only central in logic, it is also ubiquitous in linguistics and computer science. Syntax deals with the structure of a logical or linguistic expression in terms of its constituent symbols; semantics deals with mappings to objects capturing the meaning of expressions. Both are essentially about relationships between expressions, rather than about individual expressions. For instance, certain syntactic logical transformations produce new expressions from given ones by, for example, renaming or unifying variables. Semantics tells us under which conditions the syntactically modified expressions are equivalent to the original ones. Syntactic transformations can be chained together to form *derivations*, chains of expressions each of which is obtained from the previous one by one of the possible transformations. Semantics, on the other hand, is mostly concerned with the relation between the initial and final expressions. Syntax is more concerned with *how* to compute the final expression from the initial one, while semantics is more concerned with *what* the relation between them is. This what–how duality permeates all of computer science: from specification-design, via grammar-parser, to declarative-imperative programming.

A semantic relation with particular significance in mathematics and computer programming is the relation of *logical equivalence*, requiring that under no circumstance should a syntactic operation remove or add meaning. For instance, logically equivalent statements of a theorem (such as 'there exists no largest prime number' and 'there are infinitely many primes') are essentially seen as one and the same theorem. A related notion, and in fact far more useful, is the notion of *entailment* or *logical implication*. Two expressions are logically equivalent if, and only if, each entails the other. Many syntactic transformations produce weaker expressions that are entailed by, but not logically equivalent with, the original expression. For instance, we can *specialize* the expression 'there exists no largest prime number' to '4,220,851 is not the largest prime number.' Syntactic transformations which specialize expressions into weaker entailed expressions are called *sound* transformations. It may seem wasteful to throw away knowledge in this way, but logicians are often interested in *complete* sets of syntactic transformations which, when applied in every possible way, generate all possible implied expressions.

Soundness and completeness constitute the canon of mathematical logic. They allow us to reformulate mathematical knowledge into more manageable specializations about the particular topic we are interested in. They also allow us to combine several pieces of knowledge: for instance, from '4,220,851 is not the largest prime number' and '4,220,851 is a prime number' we can infer '4,220,851 is not the largest natural number.' Sound and complete transformations, or *inference rules* as they are often called, are also central in many areas of computer science, for instance when we want to prove that a particular computer program meets its specification. In all these cases the starting point (the mathematical axioms, the grammar, or the program specification) is already, in an abstract sense, complete. If a mathematical theorem embodies knowledge that was not already present in the axioms we started from, the theorem is simply wrong. In mathematics the only allowed form of reasoning is sound reasoning or *deduction*.

2 Non-Deductive Reasoning Forms

In experimental sciences, and indeed in everyday life, the overwhelming majority of inferences is not deductive. Any physical theory that is to be of any use is expected to *generalize* the observations, in the sense that it makes predictions about as yet unobserved phenomena. If inference of such a theory from observations were required to be sound, no such predictions would be possible. Similarly, if our observations are insufficient to warrant a certain conclusion, we are usually happy to make educated guesses about the missing knowledge, even if this renders our inference, strictly speaking, unsound. The good news about giving up soundness is that our inferences may become much more useful; the bad news is that they may turn out to be wrong.

The fact that in science and everyday life non-deductive reasoning is ubiquitous suggests that we humans are relatively successful in avoiding most of the pitfalls of unsound reasoning, and that our non-deductive inferences are none the less correct most of the time. It follows that unsound reasoning comes in kinds – for instance, there is a trivial distinction between incorrect reasoning (such as inferring that all swans are black after observing 10 white swans) from unsound but potentially correct reasoning (such as inferring that all swans are white from the same observations). More interestingly, we would expect there to be different forms of unsound reasoning: one to deal with missing premises, one to propose a theory generalizing given observations, one for performing what-if analysis, one to explain observed behavior of a particular object, and so on. We would also expect to have some way to assess the reliability of an unsound inference, expressed in terms of, for example, the predictions it makes, the explanations it provides, the assumptions it requires, and the observations on which it was based.

There is a plethora of interesting research questions to explore. Which different kinds of unsound reasoning can be meaningfully distinguished? How different is each of them from deduction? Can we draw up a list of necessary and sufficient conditions for any kind of reasoning to be called deductive? Can we remove conditions from this list, and still obtain sensible but unsound forms of reasoning? Are soundness and completeness relative notions, for example does it make sense to talk about *inductive* soundness as distinct from deductive soundness? All these are issues one would expect to be central on most logicians' agendas. Yet, they seem to have fallen off during the 'Golden Age':

> The central process of reasoning studied by modern logicians is the accumulative deduction, usually explained semantically, as taking us from truths to further truths. But

> actually, this emphasis is the result of a historical contraction of the agenda for the field. Up to the 1930s, many logic textbooks still treated deduction, induction, confirmation, and various further forms of reasoning in a broader sense as part of the logical core curriculum. And moving back to the 19th century, authors like Mill or Peirce included various non-deductive modes of reasoning (induction, abduction) on a par with material that we would recognize at once as 'modern' concerns. Since these non-deductive styles of reasoning seemed irrelevant to foundational research in mathematics, they moved out quietly in the Golden Age of mathematical logic. But they do remain central to a logical understanding of ordinary human cognition. These days, this older broader agenda is coming back to life, mostly under the influence of Artificial Intelligence, but now pursued by more sophisticated techniques – made available, incidentally, by advances in mathematical logic. (van Benthem 2000)

To be sure: I am not arguing that logicians stopped investigating the research issues I indicated above – on the contrary, there have been many exciting developments regarding these questions, some of which will be covered in this chapter. However, they do seem to have disappeared from the main logical agenda. I believe it is important to revive the broader logical agenda, on which mathematical logic is an important subtopic but not the only one. If anything, such a broader agenda would stimulate cross-fertilization among subtopics, something which happens too seldom nowadays:

> Some members of the traditional logic community are still very conservative in the sense that they have not even accepted non-monotonic reasoning systems as logics yet. They believe that all this excitement is transient, temporarily generated by computer science and that it will fizzle out sooner or later. They believe that we will soon be back to the old research problems, such as how many non-isomorphic models does a theory have in some inaccessible cardinal or what is the ordinal of yet another subsystem of analysis. I think this is fine for mathematical logic but not for the logic of human reasoning. There is no conflict here between the new and the old, just further evolution of the subject. (Gabbay 1994: 368, note 7)

In the remainder of this chapter I will be considering the following fundamental question: which are the main forms of reasoning that make up the logical agenda, and what are their key characteristics? Informally, *reasoning* is the process of forming arguments, that is drawing conclusions from premises. By fixing the relation between premises and acceptable conclusions we can obtain various *reasoning forms*. For instance, an argument is *deductive* if the conclusion cannot be contradicted (or *defeated*) by new knowledge without contradicting the premises also; a form of reasoning is deductive if it only allows deductive arguments. We also say that deductive reasoning is *non-defeasible*. A *logical system*, or *logic* for short, is a particular formalization of a reasoning form. There may exist several logics formalizing a particular reasoning form; for instance, there is a range of deductive logics, such as modal, temporal, relevance, and intuitionistic logics, each formalizing certain aspects of deductive reasoning. These deductive logics do not necessarily agree on which arguments are deductively valid and which are not. For example, the argument 'two plus two equals four; therefore, if the moon is made of green cheese, then two plus two equals four' will be rejected by those who favor a causal or relevance interpretation of if–then rather than a truth-functional interpretation.

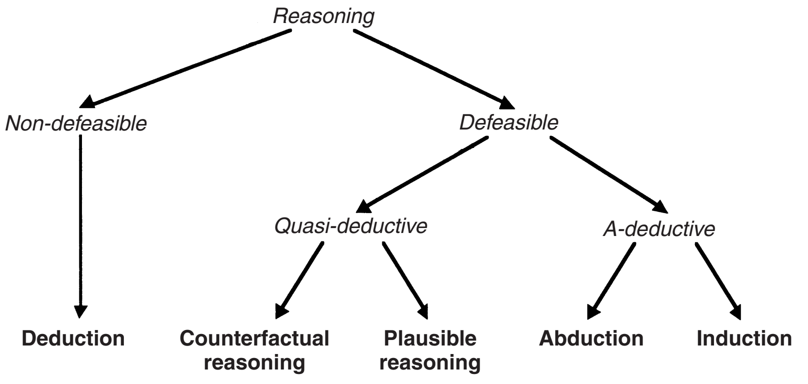

Figure 42.1 A classification of reasoning forms

However, as soon as such an argument is accepted as deductively valid, the only way to defeat the conclusion is by denying that two plus two equals four, and this defeats the premises also.

Non-deductive reasoning forms, on the other hand, are defeasible: a conclusion may be defeated by new knowledge, even if the premises on which the conclusion was based are not defeated. For instance, the argument 'birds typically fly; Tweety is a bird; therefore, Tweety flies' is non-deductive, since Tweety might be an ostrich, hence non-typical. The argument 'every day during my life the sun rose; I don't know of any trustworthy report of the sun not rising one day in the past; therefore, the sun will rise every future day' is non-deductive, since if the sun would not rise tomorrow, this would invalidate the conclusion but not the premises. The Tweety-argument is a well-known example of what I call *plausible reasoning*: reasoning with general cases and exceptions. An important observation is that plausible reasoning encompasses deductive reasoning: if we know that Tweety is a typical bird, the argument will be deductively valid. In this sense plausible reasoning is 'supra-deductive' or, as I will call it, *quasi-deductive*. Another example of quasi-deductive reasoning is so-called *counterfactual reasoning*, or 'what-if' analysis, starting from premises known to be false. For instance, the argument 'if you hadn't called me this morning, I would surely have missed my train' is a counterfactual argument, as both premise and conclusion are false in the intended interpretation. The point of such an argument is to investigate what would change if certain circumstances in the world had been different.

Other reasoning forms do not aim at approximating deduction, hence do not include deduction as a special case. I will call such reasoning forms *a-deductive*. The sunrise argument is an example of *induction*, an a-deductive reasoning form aimed at generalizing specific *observations* (also called evidence) into general rules or *hypotheses*. Note that I do not yet claim to have *defined* plausible or inductive reasoning in any way. Like with all forms of reasoning, this requires a formal definition of the consequence relation between premises and acceptable conclusions, analogous to deductive entailment. (The general term I will use for such a relation is *consequence*: thus, we will speak about 'inductive consequence' or 'plausible consequence,' and avoid potentially confusing terms like 'inductive validity' or 'plausible soundness').

Another form of a-deductive reasoning is *abduction*, a term originally introduced by C. S. Peirce to denote the process of forming an explanatory hypothesis given some observations (a hypothesis from which the observations can be deduced). For instance, the argument 'All the beans from this bag are white; these beans are white; therefore, these beans are from this bag' is an abductive argument. In recent years, abduction has become popular in the logic programming field, where it denotes a form of reasoning where the general explanation is known, but one of its premises is not known to be true; abduction is then seen as hypothesizing this missing premise. As a consequence, abduction and induction are viewed as complementary: induction infers the general rule, given that its premises and its conclusion hold in specific cases; abduction infers specific premises, given the general rule, and specific instances of its conclusion and some of its premises. Also, there are strong links between abduction and plausible reasoning: abduction can answer the question 'what do I need to assume about the bird Tweety if I want to infer that it flies' (answer: that it is a typical bird). I will expand on some of these issues below – the reader interested in finding out more about the relation between abduction and induction is referred to (Flach and Kakas 2000a).

The classification of reasoning forms I am advocating is depicted in Figure 42.1. While the justification for some of the distinctions made here have been admittedly sketchy, they will be elaborated in the rest of the chapter. The reader should also be aware that this classification should be taken as a starting point and is not intended to be set in stone. The main point is that on the map of logic, deduction occupies but a small part. I will now proceed to discuss some of these reasoning forms in more detail.

3 Plausible Reasoning

I should start by stressing that the term 'plausible reasoning' is not generally accepted – reasoning with exceptions is normally referred to as non-monotonic reasoning. *Monotonicity* is a technical term denoting that the set of conclusions grows (monotonically) with the set of premises. In other words, addition of a premise to a given argument never invalidates the conclusion – the same property as what I called non-defeasibility above. Since any non-deductive reasoning form is defeasible, it follows that any non-deductive reasoning form is non-monotonic. Thus, the property of non-monotonicity is of limited use in singling out a particular non-deductive reasoning form; for this reason I prefer a different (and more meaningful) term for reasoning with general rules and exceptions. (*Default reasoning* would be a good term, but this seems too strongly connected to a particular logic, i.e. default logic.)

Plausible reasoning is the process of 'tentatively inferring from given information rather more than is deductively implied' (Makinson 1994). It can thus be said to be more liberal or more credulous than deductive reasoning. Correspondingly, the set of arguments accepted by a plausible reasoning agent (also called a *consequence relation*, and defined as a subset of $L \times L$, where L is the language) can be divided into a deductive part and a plausible part. The deductive part corresponds to arguments not involving any rules which have exceptions. (Alternatively, one can deductively extend a set of plausible arguments by treating all exceptions to rules as inconsistencies from which everything could be inferred, although this would be rather less interesting.)

The non-monotonicy of plausible reasoning can be demonstrated as follows: from *bird* one would infer *flies*, but from *bird and penguin* one wouldn't infer *flies*. That is, the rule *if bird then flies* is a *default rule* which tolerates exceptions, and the formula *bird and not flies* is not treated as an unsatisfiable formula from which anything can be inferred. The question then arises as to what other properties of deductive reasoning, besides monotonicity, are affected by allowing exceptions to rules. This is the main question addressed in a seminal paper by Kraus et al. (1990). In general, propositional deductive reasoning can be characterized by the following rules:

Reflexivity: $\alpha \vdash \alpha$ for all α;
Monotonicity: if $\alpha \vdash \beta$ and $\gamma \models \alpha$, then $\gamma \vdash \beta$;
Right Weakening: if $\alpha \vdash \beta$ and $\beta \models \gamma$, then $\alpha \vdash \gamma$;
Cut: if $\alpha \vdash \beta$ and $\alpha \wedge \beta \vdash \gamma$, then $\alpha \vdash \gamma$;
Left Or: if $\alpha \vdash \gamma$ and $\beta \vdash \gamma$, then $\alpha \vee \beta \vdash \gamma$.

In these rules, $\alpha \vdash \beta$ indicates that the reasoner in question accepts the inference from α to β, possibly with respect to an implicit body of background knowledge. $\models$, on the other hand, stands for classical deductive consequence (with respect to the same background knowledge). These rules can be combined: for instance, Reflexivity and Right Weakening together imply that $\alpha \vdash \gamma$ whenever $\alpha \models \gamma$, that is the consequence relation $\vdash$ is *supra-classical*.

Kraus et al. prove that the above five rules characterize deductive reasoning. Notice that equivalent rule sets exist: for instance, Cut could be replaced by Right And, and Left Or could be replaced by Right Implication:

Right And: if $\alpha \vdash \beta$ and $\alpha \vdash \gamma$, then $\alpha \vdash \beta \wedge \gamma$;
Right Implication: if $\alpha \wedge \beta \vdash \gamma$, then $\alpha \vdash \beta \rightarrow \gamma$.

Furthermore, they study the kinds of reasoning that result from weakening some of these rules. One variant they consider is obtained by replacing Monotonicity with the following two rules:

Left Logical Equivalence: if $\alpha \vdash \beta$ and $\models \alpha \leftrightarrow \gamma$, then $\gamma \vdash \beta$;
Cautious Monotonicity: if $\alpha \vdash \beta$ and $\alpha \vdash \gamma$, then $\alpha \wedge \beta \vdash \gamma$.

Both rules are clearly entailed by Monotonicity – Cautious Monotonicity, in particular, states that premises can be strengthened with their plausible consequences. This kind of plausible reasoning is called *preferential* reasoning, because it can be semantically modeled by assuming a (partial) preference order between states, where a state is a set of models, and stipulating that $\alpha \vdash \beta$ if and only if every *most preferred* state satisfying α also satisfies β (a state satisfies a formula iff all its models satisfy the formula). Preferential reasoning can be further weakened by dropping the condition that the preference relation between states be a partial order; this invalidates Left Or but none of the other rules. This kind of reasoning is called *cumulative* reasoning because Cut and Cautious Monotonicity together imply that if $\alpha \mid\!\sim \beta$, then $\alpha \mid\!\sim \gamma$ if and only if $\alpha \wedge \beta \mid\!\sim \gamma$, that is plausible consequences can be accumulated in the premises.

From the foregoing it follows that deductive reasoners are preferential (with empty preference relation), and preferential reasoners are cumulative. However, a more meaningful comparison between reasoning forms *X* and *Y* would be obtained if we could establish, for each *X*-reasoner, a unique maximal subset of arguments that satisfy the rules of *Y*. Such a *reduction* from preferential to deductive reasoning was given in (Flach 1998). Basically, it involves using Monotonicity in the opposite direction (if $\gamma \models \alpha$ and $\gamma \not\models \beta$, then $\alpha \not\models \beta$) to remove arguments that are not deductively justified. Semantically, this amounts to ignoring the preference relation. (As stated before, we can also use Monotonicity in the forward direction to turn all plausible arguments into deductive ones, amounting to removing all states that satisfy exceptions to rules; however, this would rather endorse the less natural view that plausible reasoning is the process of inferring *less* than deductively implied.)

There is an interesting analogy between non-monotonic reasoning and non-Euclidean geometry. For many centuries it was assumed that Euclid's fifth axiom (parallel lines don't intersect) was self-evident, and that denying it would lead to inconsistencies. However, non-Euclidean geometry was proved to be consistent in the early nineteenth century. Similarly, many logicians argued that logic was necessarily monotonic, and that the concept of a non-monotonic logic was a contradiction in terms. However, there is a difference between monotonicity as a property of mathematical reasoning, and monotonicity of the logic under study. Kraus et al. used the deductive metalogic of consequence relations to formalize various forms of non-deductive reasoning. Rules such as Cautious Monotonicity are in fact *rationality postulates* that need to be satisfied by any rational reasoning agent of the class under study. This is a crucial insight, and their approach establishes a methodology that can be applied to analyze other forms of reasoning as well. This will be explored in the next section.

4 Induction and Abduction

Induction is the process of generalizing specific evidence into general rules. A simple form of induction is the following sample-to-population inference:

X percent of observed *F*s are *G*s;
therefore, (approximately) *X* percent of all *F*s are *G*s.

This argument schema has a categorical counterpart:

All of observed *F*s are *G*s;
therefore, all *F*s are *G*s.

or – since the induced rule need not be a material implication –

All objects in the sample satisfy $P(x)$;
therefore, all objects in the population satisfy $P(x)$.

These formulations of inductive generalization, however, obscure a crucial issue: normally, the predicate *P* to be used in the general rule is not explicitly given in the observations. Rather, the key step in induction is to distill, out of all the available information about the sample, the property that is common to all objects in the sample and that will generalize reliably to the population. I will refer to this step as *hypothesis generation*.

Hypothesis generation is an often ignored step in philosophy of science. For instance, in *Conjectures and Refutations* Popper describes at length how to test a conjecture, but remains silent about how to come up with a conjecture in the first place. To refer to ill-understood phenomena such as creativity in this context is to define the problem away. Moreover, if we want to automate scientific discovery or learning (object of study in the subfield of artificial intelligence called machine learning), we have to approach hypothesis generation in a principled way. Hypothesis generation is not a wholly irrational process, and the question thus becomes: what are the rationality postulates governing inductive hypothesis generation?

In fact, this question was already considered by the American philosopher Charles Sanders Peirce, who wrote in 1903:

> 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." (Peirce 1958: 5.188–9)

Here, Peirce calls the process of explanatory hypothesis generation *abduction* (while he uses the less tentative phrase "adopting an explanatory hypothesis" above, elsewhere (5.171) he defines abduction as "the process of forming an explanatory hypothesis," i.e. "abduction merely suggests that something *may be*").

Nowadays people use the term 'abduction' in various senses (even Peirce had initially a different, syllogistic view of abduction), so a brief digression on these issues may be in order – the interested reader is referred to (Flach and Kakas 2000b) for a more extensive discussion. In philosophy, it is customary to view abduction as 'reasoning to the best explanation' (Lipton 1991). This, however, combines hypothesis generation with hypothesis selection, only the former being a purely logical process amenable to logical analysis. In artificial intelligence, abduction is usually perceived as reasoning from effects to causes, or from observations to explanations: here, an abductive hypothesis is not a general rule or theory, as in induction, but rather a specific explanation or cause relating to the observed individual. Thus, abductive hypotheses explain but do not generalize. Induction, on the other hand, aims at generalizing beyond the observed individuals. While in inductive argument schemas such as the above the induced hypotheses entails the observations, this is not an explanation in the same sense as a cause explains an effect.

In general, we cannot distinguish between abductive explanations and inductive generalisations by methods based on entailment alone, including the method I am about to describe. However, the view of induction as generalization does suggest an alternative formalization which is closer to both confirmation theory (in the qualitative sense of Hempel) and Kraus et al.'s (1990) account of plausible reasoning. In the remainder of this section I will discuss rationality postulates for explanatory reasoning, including abduction and explanatory induction, and then present alternative postulates for confirmatory induction in the next section.

Returning to Peirce, the logical form of abductive hypothesis generation he suggested can be simplified to 'from C, and $A \models C$, abduce A' or, introducing the symbol $\mathrel{|\!<}$ for abductive inference, 'if $A \models C$, then $C \mathrel{|\!<} A$'. We can now use Kraus et al.'s (1990) consequence relation methodology to formulate rationality postulates for hypothesis generation. We start with some general principles:

Verification: if $\alpha \mathrel{|\!<} \beta$ and $\alpha\wedge\beta \models \gamma$, then $\alpha\wedge\gamma \mathrel{|\!<} \beta$;
Falsification: if $\alpha \mathrel{|\!<} \beta$ and $\alpha\wedge\beta \models \gamma$, then $\alpha\wedge\neg\gamma \not\mathrel{|\!<} \beta$.

Verification and Falsification state that if β is a possible hypothesis given observations α, and γ is a prediction on the basis of β (and α), then β is not ruled out by observing that γ is true, but falsified by observing that γ is false. (While the names of these rules have been inspired by the debate between the logical positivists and Popper, it should be stressed that – under my interpretation of $\alpha \mathrel{|\!<} \beta$ as 'β is a *possible* hypothesis given evidence α' – Verification is a fairly weak rule to which one can hardly object.)

Falsification is different from the rules we have seen until now, because it draws negative conclusions about the consequence relation $\mathrel{|\!<}$. This means that some of Kraus et al.'s (1990) rules need to be adapted when formulated in our framework. For instance, the following set of 'explanatory' rules is obtained by rewriting the rules given in Section 3 for deduction, substituting $\beta \mathrel{|\!<} \alpha$ for $\alpha \mathrel{|\!\sim} \beta$ (we use the variant with Right And and Right Implication):

Reflexivity: $\alpha \mathrel{|\!<} \alpha$ for all α;
Right Strengthening: if $\beta \mathrel{|\!<} \gamma$ and $\alpha \models \gamma$, then $\beta \mathrel{|\!<} \alpha$;
Left Weakening: if $\beta \mathrel{|\!<} \alpha$ and $\beta \models \gamma$, then $\gamma \mathrel{|\!<} \alpha$;
Left And: if $\beta \mathrel{|\!<} \alpha$ and $\gamma \mathrel{|\!<} \alpha$, then $\beta\wedge\gamma \mathrel{|\!<} \alpha$;
Left Implication: if $\gamma \mathrel{|\!<} \alpha\wedge\beta$, then $\beta\rightarrow\gamma \mathrel{|\!<} \alpha$.

The last three rules make immediate sense for explanatory hypothesis generation. In particular, Left Weakening states that the set of explanations decreases monotonically when the observations increase; it is a convergence property for induction (it can be combined with Verification into a single rule). Left And states that if α is a possible hypothesis explaining β and γ observed separately, it also explains β and γ observed together; this enables incremental induction. Left Implication deals with background knowledge: if β is a necessary part of the explanation of γ, then it can also be added as a condition to the observation.

On the other hand, the first two rules contradict Falsification and need to be weakened by adding an admissibility requirement on α (for instance, that α explains something – e.g. itself). Without going into details, we mention that the following set of rules has been demonstrated to characterize consistent explanatory reasoning:

Explanatory Reflexivity: if $\beta \mathrel{|\!<} \beta$ and $\neg\alpha \not\mathrel{|\!<} \beta$, then $\alpha \mathrel{|\!<} \alpha$;
Admissible Right Strengthening: if $\beta \mathrel{|\!<} \gamma$, $\alpha \mathrel{|\!<} \alpha$ and $\alpha \models \gamma$, then $\beta \mathrel{|\!<} \alpha$;
Predictive Left Weakening: if $\beta \mathrel{|\!<} \alpha$ and $\alpha\wedge\beta \models \gamma$, then $\gamma \mathrel{|\!<} \alpha$;
Left And: if $\beta \mathrel{|\!<} \alpha$ and $\gamma \mathrel{|\!<} \alpha$, then $\beta\wedge\gamma \mathrel{|\!<} \alpha$;
Left Implication: if $\gamma \mathrel{|\!<} \alpha\wedge\beta$, then $\beta\rightarrow\gamma \mathrel{|\!<} \alpha$;
Left Consistency: if $\alpha \mathrel{|\!<} \beta$ then $\neg\alpha \not\mathrel{|\!<} \beta$.

While some of these postulates may be debatable (for instance, one may argue that explanatory reasoning is inherently irreflexive), they do provide a starting point for studying various forms of explanatory reasoning. Instead of a single 'logic of induction', I have proposed a modular system of meta-level rationality postulates that can be adapted to model various forms of reasoning. In addition, one can study semantic characterizations of these postulates. The interested reader is referred to (Flach 2000a, 2000b).

5 Confirmatory Induction

The preceding set of postulates concentrated on induction and abduction as explanatory reasoning. There is an alternative view of induction as inferring hypotheses that are *confirmed* by the observations. This view was pioneered by Carl G. Hempel, who proposed both a set of rationality postulates (or, as he called them, adequacy conditions) and a material definition of confirmation. The following is a list of Hempel's adequacy conditions (Hempel, 1945: 103–6, 110), reformulated in our meta-language:

Entailment: if $\alpha \models \beta$, then $\alpha \mathrel{|\!<} \beta$;
Right Weakening: if $\alpha \mathrel{|\!<} \beta$ and $\beta \models \gamma$, then $\alpha \mathrel{|\!<} \gamma$;
Right And: if $\alpha \mathrel{|\!<} \beta$ and $\alpha \mathrel{|\!<} \gamma$, then $\alpha \mathrel{|\!<} \beta\wedge\gamma$;
Consistency: if $\alpha \mathrel{|\!<} \beta$ and $\alpha \not\models \neg\alpha$, then $\alpha \not\models \neg\beta$;
Left Logical Equivalence: if $\alpha \mathrel{|\!<} \beta$ and $\models \alpha\leftrightarrow\gamma$, then $\gamma \mathrel{|\!<} \beta$;

For instance, the first condition states that entailment 'might be referred to as the special case of conclusive confirmation' (Hempel 1945: 107). Each of these postulates is reasonable, except perhaps Right And which seems unjustified if the evidence is too weak to rule out incompatible hypotheses – in other words, it expresses a completeness assumption regarding the observations.

The main reason for Hempel to formulate his adequacy conditions was to verify his material definition of confirmation against them – consequently, there is no guarantee that they are complete in any sense. The following set of rationality postulates for confirmatory induction can be shown to be complete with respect to a suitably devised semantics:

Confirmatory Reflexivity: if $\beta \mathrel{|\!<} \beta$ and $\beta \not\mathrel{|\!<} \neg\alpha$, then $\alpha \mathrel{|\!<} \alpha$;
Predictive Right Weakening: if $\alpha \mathrel{|\!<} \beta$ and $\alpha\wedge\beta \models \gamma$, then $\alpha \mathrel{|\!<} \gamma$;

Right And: if $\alpha \mathrel{|\!<} \beta$ and $\alpha \mathrel{|\!<} \gamma$, then $\alpha \mathrel{|\!<} \beta\wedge\gamma$;
Right Consistency: if $\alpha \mathrel{|\!<} \beta$ then $\alpha \not\mathrel{|\!<} \neg\beta$;
Left Logical Equivalence: if $\alpha \mathrel{|\!<} \beta$ and $\models \alpha\leftrightarrow\gamma$, then $\gamma \mathrel{|\!<} \beta$;
Strong Verification: if $\alpha \mathrel{|\!<} \beta$ and $\alpha \mathrel{|\!<} \gamma$, then $\alpha\wedge\gamma \mathrel{|\!<} \beta$;
Left Or: if $\alpha \mathrel{|\!<} \gamma$ and $\beta \mathrel{|\!<} \gamma$, then $\alpha\vee\beta \mathrel{|\!<} \gamma$.

As before, I disallow contradictory observations (unlike Hempel) – a weaker form of Entailment follows from Predictive Right Weakening, and a weak form of Reflexivity has been added as a separate rule (notice that Reflexivity was implied by Hempel's rules as an instance of Entailment). Two new rules have been added. Whereas Verification states that predictions γ can be added to confirming observations α for hypothesis β, Strong Verification states that this can also be done whenever γ is confirmed by α. As with Right And, the underlying assumption is that the observations are complete enough to have all confirmations 'point in the same direction.' Left Or can be seen as a variant of Left Weakening, discussed in the context of explanatory reasoning. While Left Weakening is clearly invalid in the confirmatory case (if we weaken the observations, there will presumably come a point where they cease to confirm the hypothesis), Left Or states that separate observations confirming a hypothesis can be weakened by taking their disjunction.

The semantics against which these postulates are provably complete is a variant of Kraus et al.'s (1990) preferential semantics for plausible reasoning. In fact, the postulates for confirmatory induction are closely related to postulates considered in Section 3: for instance, Strong Verification is identical with Cautious Monotonicity. This is perhaps surprising at first sight, but can be explained by noting that plausible and confirmatory reasoning make similar assumptions in order to go beyond deduction: while in plausible reasoning one commonly assumes that anything which is not known to be an exception conforms to the rule, in induction one assumes that unknown objects behave similarly to known objects.

We end this section on a philosophical note. Hempel's name is associated with a number of paradoxes, one of which is the *confirmation paradox*. This paradox arises when one considers to add a variant of Right Strengthening to the postulates for confirmatory induction. To borrow Hempel's example:

> Is it not true, for example, that those experimental findings which confirm Galileo's law, or Kepler's laws, are considered also as confirmation Newton's law of gravitation?' (Hempel 1945: 104)

The problem is that the combination of Right Weakening and Right Strengthening immediately leads to a collapse of the system, since arbitrary observations now confirm arbitrary hypotheses. However, Hempel confuses confirmation with explanation here. Explanatory hypotheses can be arbitrarily strengthened (as long as they remain consistent with the observations), but not necessarily weakened; confirmed hypotheses can be arbitrarily weakened, but only strengthened under certain conditions. It might be possible to formalize a form of hypothesis generation where hypotheses both explain and are confirmed by the observations (this is an open problem), but then there would be strong conditions on both strengthening and weakening of the hypothesis. Distinguishing between explanatory and confirmatory induction solves the confirmation paradox.

6 Concluding Remarks

This short chapter has been written as a re-appraisal of logic as the science of knowledge. The goal of logic is to provide a catalogue of reasoning forms. Deduction is but one of the possible forms of reasoning, easiest to formalize but with limited importance for intelligence. It is possible to characterize non-deductive or defeasible reasoning forms mathematically, by concentrating on their purely logical part, *viz*. hypothesis generation. I have suggested that such characterization is best performed on the meta-level, stating postulates that circumscribe rational behavior of reasoning agents. Possible rationality postulates for plausible, explanatory, and confirmatory reasoning have been discussed at some length.

A final word on the issue of hypothesis *selection*, which is the equally crucial but complementary step in intelligent reasoning. In my view, the process of evaluating possible hypotheses to determine which one(s) will be actually adopted is an extra-logical one. By this I mean that it does not give rise to a proof theory in any interesting sense. Furthermore, any hypothesis evaluation procedure will be construed from measures of probability, interestingness, or information content. Logic deals with possible conclusions, not actual ones. This is even true for deduction, which only characterizes tautologies, not interesting mathematical theorems. My conjecture is that successful evaluation procedures (e.g. based on Bayesian or subjective probabilities) will be applicable across a range of different reasoning forms. Thus, while hypothesis generation distinguishes reasoning forms, hypothesis evaluation unifies them.

References

van Benthem, J. (2000) Reasoning in reverse. In Flach and Kakas (2000a): ix–xi.

Flach, P. A. (1998) Comparing consequence relations. In A. G. Cohn, L. Schubert and S. C. Shapiro (eds.), *Proceedings of the Sixth International Conference on Principles of Knowledge Representation and Reasoning*. Morgan Kaufmann: 180–9.

Flach, P. A. (2000a) On the logic of hypothesis generation. In Flach and Kakas (2000a): 89–106.

Flach, P. A. (2000b) Logical characterisations of inductive learning. In D. M. Gabbay and P. Smets (eds.), *Handbook of Defeasible Reasoning and Uncertainty Management*, vol. 4: *Abduction and Learning*. Dordrecht: Kluwer Academic.

Flach, P. A. and Kakas, A. C. (eds.) (2000a) *Abduction and Induction: Essays on their Relation and Integration*. Dordrecht: Kluwer Academic.

Flach, P. A. and Kakas, A. C. (2000b) Abductive and inductive reasoning: background and issues. In Flach and Kakas (2000a): 1–27.

Gabbay, D. M. (1994) Classical vs non-classical logics (the universality of classical logic). In D. M. Gabbay, C. J. Hogger and J. A. Robinson (eds.), *Handbook of Logic in Artificial Intelligence and Logic Programming*, vol. 2: *Deduction Methodologies*. Oxford: Clarendon Press, 359–495.

Hempel, C. G. (1945) Studies in the logic of confirmation. *Mind*, 54(213 & 214), 1–26 and 97–121.

Kraus, S., Lehmann, D. and Magidor, M. (1990) Nonmonotonic reasoning, preferential models and cumulative logics. *Artificial Intelligence*, 44, 167–207.

Makinson, D. (1994) General patterns in nonmonotonic reasoning. In *Handbook of Logic in Artificial Intelligence and Logic Programming,* vol. 3. Oxford: Clarendon Press, 35–110.

Lipton, P. (1991) *Reasoning to the Best Explanation*. London: Routledge and Kegan Paul.

Peirce, C. S. (1958) C. Harstshorne, P. Weiss and A. Burks (eds.), *Collected Papers of Charles Sanders Peirce*. Cambridge, MA: Harvard University Press.