ABDUCTIVE REASONING

LOGICAL INVESTIGATIONS INTO DISCOVERY AND EXPLANATION

by

ATOCHA ALISEDA
National Autonomous University of Mexico

Springer

Chapter 3

ABDUCTION AS LOGICAL INFERENCE

1. Introduction

In the preceding overview chapter, we have seen how the notion of abduction arose in the last century out of philosophical reflection on the nature of human reasoning, as it interacts with patterns of explanation and discovery. Our analysis brought out a number of salient aspects to the abductive process, which we shall elaborate in a number of successive chapters. For a start, abduction may be viewed as a kind of logical inference and that is how we will approach it in the analysis to follow here. Evidently, though, as we have already pointed out, it is not standard logical inference, and that for a number of reasons. Intuitively, abduction runs in a *backward* direction, rather than the *forward* one of standard inference, and moreover, being subject to revision, it exhibits non-standard non-monotonic features (abductive conclusions may have to be retracted in the light of further evidence), that are more familiar from the literature on non-standard forms of reasoning in artificial intelligence. Therefore, we will discuss abduction as a broader notion of consequence in the latter sense, using some general methods that have been developed already for non-monotonic and dynamic logics, such as systematic classification in terms of structural rules. This is not a mere technical convenience. Describing abduction in an abstract general way makes it comparable to better-studied styles of inference, thereby increasing our understanding of its contribution to the wider area of what may be called 'natural reasoning'. To be sure, in this chapter we propose a logical characterization of what we have called an *(abductive) explanatory argument*, in order to make explicit that the inference is explanatory (and thus forward chained), while keeping in mind it aims to characterize the conditions for an abductive explanation to be part of this inference (cf. chapter 2).

The outcomes that we obtain in this first systematic chapter, naturally divided into five parts, are as follows. After this introduction, in the second part (section 2), we discuss the problem of demarcation in logic, in order to set the ground for our analysis of abduction as a logical inference. Placing abduction in a broader universe of logics and stances, make natural to consider it as a logical inference of its own kind. In the third part (section 3), we discuss in detail some aspects of abductive inference, such as its direction, format of premisses and conclusion as well as its inferential strength. We propose a general logical format for abduction, involving more parameters than in standard inference, allowing for genuinely different roles of premisses. We find a number of natural styles of abductive explanatory inference, rather than one single candidate. These (abductive) explanatory versions are classified by different *structural rules* of inference, and this issue occupies the fourth part (section 4). As a contribution to the logical literature in the field, we give a complete characterization of one simple style of abductive explanatory inference, which may also be viewed as the first structural characterization of a natural style of explanation in the philosophy of science. We then analyze some other abductive explanatory versions (explanatory, minimal and preferential) with respect to their structural behaviour, giving place to more sophisticated structural rules with interest of their own. Finally, we turn to discuss further logical issues such as how those representations are related to more familiar completeness theorems, and finally, we show how abduction tends to involve higher complexity than classical logic: we stand to gain more explanatory power than what is provided by standard inference, but this surplus comes at a price. In the fifth and final part of this chapter (section 5),we offer an analysis of previous sections centering the discussion on abduction as an enriched form of logical inference with an structure of its own. We then put forward our conclusions and present related work within the study of non-monotonic reasoning.

Despite these useful insights, pure logical analysis does not exhaust all there is to abduction. In particular, it's more dynamic process aspects, and its interaction with broader conceptual change must be left for subsequent chapters, that will deal with computational aspects, as well as further connections with the philosophy of science and artificial intelligence. What we do claim, however, is that our logical analysis provides a systematic framework for studying the intuitive notion of abductive explanatory inference, which gives us a view of its variety and complexity, and which allows us to raise some interesting new questions.

2. Logic: The Problem of Demarcation

One of the main questions for logic is the problem of demarcation, that is, to distinguish between logical and non-logical systems. This question is at the core of the philosophy of logic, and has a central place in the philosophy of mathematics, in the philosophy of science as well as in the foundations of artificial intelligence.

Some questions in need for an answer for this problem concern the following ones: what is a logic?, which is the scope of logic?, which formal systems qualify as logics?, all of these leading to metaphysical questions concerning the notion of correctness of a logical system: does it make sense to speak of a logical system as correct or incorrect?, could there be several logical systems which are equally correct?, is there just one correct logical system? These questions in turn lead to epistemological questions of the following kind: how does one recognise a truth of logic? could one be mistaken in what one takes to be such truths?

There are however, several proposals and positions in the literature in regard to all these questions. Our strategy to describe the problem of demarcation of logic will be the following. Our point of departure is Peirce's distinction of three types of reasoning, namely deduction, induction, and abduction. We will compare them according to their certainty level, something that in turn gives place to different areas of application, mainly in mathematics, philosophy of science and artificial intelligence. Next, we will introduce an standard approach in philosophy of logic based on the relationship between informal arguments and their counterparts in formal logic, namely the view endorsed by Haack [Haa78]. Her classification of kinds of logics will be presented, that is, the well-known distinction amongst extensions and deviations of classical logics, and inductive logics. Moreover, we take up on Haack's discussion on the several positions with respect to the legitimization (and proliferation) of logics, namely instrumentalism, monism and pluralism. Finally, we will introduce a much less-known approach –but still standard within its field – coming from artificial intelligence, namely the logical structural approach devised for the study of non-monotonic reasoning.

Our overall discussion in this section will serve two purposes. On the one hand, it aims to show that even under a broad view of logic, there is neither a unique nor a definite answer to the problem of demarcation, not to mention to each of the former questions. On the other hand, it will set the ground for the main purpose of this chapter, that is, an analysis of abduction as a specific kind of logical inference, in order to show, that abduction holds a natural place to be considered a logical inference of its own kind.

Types of Reasoning: Deduction, Induction, Abduction

From a logical perspective, mathematical reasoning may be identified with classical, deductive inference. Two aspects are characteristic of this type of reasoning, namely its *certainty* and its *monotonicity*. The first of these is exemplified by the fact that the relationship between premisses and conclusion is that of necessity; a conclusion drawn from a set of premisses, necessarily follows from them. The second aspect states that conclusions reached via deductive reasoning are non-defeasible. That is, once a theorem has been proved, there is no doubt of its validity regardless of further addition of axioms and theorems to the system.

There are however, several other types of formal non-classical reasoning, which albeit their lack of complete certainty and monotonicity, are nevertheless rigorous forms of reasoning with logical properties of their own. Such is the case of inductive and abductive reasoning. As a first approximation, Charles S. Peirce distinction seems useful. According to him, there are three basic types of logical reasoning: *deduction*, *induction* and *abduction*. Concerning their certainty level: *'Deduction proves that something must be; Induction shows that something actually is operative; Abduction merely suggests that something may be'* [CP, 5.171]. Therefore, while deductive reasoning is completely certain, inductive and abductive reasoning are not. *'Deduction is the only necessary reasoning. It is the reasoning of mathematics'* [CP, 4.145]. Induction must be validated empirically with tests and experiments, therefore it is defeasible; and abductive reasoning can only offer hypotheses that may be refuted with additional information. For example, a generalization reached by induction (e.g. all birds fly), remains no longer valid after the addition of a premisse, which refutes the conclusion (e.g. penguins are birds). As for abduction, a hypothesis (e.g. it rained last night) which explains an observation (e.g. the lawn is wet), may be refuted when additional information is incorporated into our knowledge base (e.g. it is a drought period).

Deductive reasoning has been the paradigm of mathematical reasoning, and its logic is clearly identified with Tarski's notion of logical inference. In contrast, inductive and abductive types of reasoning are paradigmatic types of reasoning in areas like philosophy of science, and more recently, artificial intelligence. Regarding the former, contemporary research indicates that many questions regarding their logic remain controversial. As it is well known, Carnap's proposal for an inductive logic[Car55] found ample criticisms. As for abduction, while some scholars argue that the process of forming an explanatory hypothesis (our abductive process) cannot be logically reconstructed [Pop59, Hem65], and have instead proposed each a logical characterization of explanation (our (abductive) explanatory argument)[1]; others have tried to formally characterize

[1] As for the roots and similarities of these two models of explanation, Niiniluoto[Nii00, p. 140] rightly observes: *"After Hempel's (1942) paper about the deductive–nomological pattern of historical explanation, Karl Popper complained that Hempel had only reproduce his theory of causal explanation, originally presented in 'Logik der Forschung' (1935, see Popper 1945, chap 25, n. 7; Popper 1957, p. 144). With his charming politeness, Hempel pointed out that his account of D–N explanation is 'by no means novel' but 'merely summarizes and states explicitly some fundamental points which have been recognized by many scientists and methodologists".*

'retroduction' (another term for abduction), as a form of inversed deduction [Han61], but no acceptable formulation has been found. Regarding the latter, recent logico-computational oriented research has focus on studying non-standard forms of reasoning, in order to build computer programs modeling human reasoning, which being subject to revision, is uncertain and exhibits non-standard non-monotonic features. Several contemporary authors propose a more finely structured algorithmic description of logics. This concern is found both in the logical tradition ([Gab94a, vBe90]), as well as in work in philosophy ([Gil96]).

Logics: Extensions, Deviations, Inductive

Haack[Haa78] takes as primitive an intuitive notion of a formal system, and from there it hints at the characterization of what is to be a logical system, as follows:

> *"The claim for a formal system to be a logic depends, I think, upon its having an interpretation according to which it can be seen as aspiring to embody canons of valid argument."* [Haa78, page 3].

The next problem to face is that of deciding what counts as valid argumentation. But before we get into her own answer to this question, here are other criteria aiming to characterize what counts as a logical system. On the one hand, according to Kneale, logical systems are those that are purely formal, for him, those that are *complete* (in which all universal valid formula are theorems). According to Dummet, on the other hand, logical systems are those which characterize *precise* notions. Following the first characterization, many formal systems are left out, such as second order logic. If we follow the second one, then proposals such as Hintikka's system of epistemic logic is left out as well, for the notions of knowledge and belief characterize pretty vague epistemic concepts [Haa78, page 7]. Both these characterizations provide purely formal criteria for logical demarcation. For Haack, however *'the prospects for a well-motivated formal criterion are not very promising'*[Haa78, page 7], for it has the drawback of limiting the scope of logic to the point of even discarding well accepted formal systems (e.g. predicate logic) on the basis of being in absence of other metalogical properties (e.g. decidability). Moreover, many logical systems are indeed undecidable, incomplete, but nevertheless have interesting applications and have proved useful in areas like computer science and linguistics.

Haack takes a broad view of logic, considering that *'the demarcation is not based on any very profound ideas about 'the essential nature of logic'*[Haa78, page 4], and follows 'the benefit of the doubt policy', according to which, arguments may be assessed by different standards of validity, and thus accepts several formal systems as logical. For her, the question we should be asking is whether a system is good and useful rather than 'logical', which after all is not a well-defined concept. Her approach however, is not wholly arbitrary, for it does not give up the requirement of being rigorous, and takes classical logic as its reference point, building up a classification of systems of logic based on analogies to the classical system, as follows:

- Extensions (*e*)

  Modal, Epistemic, Eroretic, ...

- Deviations (*d*)

  Intuitionistic, Quantum, Many-valued, ...

- Inductive (*i*)

  Inductive probability logic

Extensions (*e*) are formal logical systems, which extend the system of classical logic ($\mathcal{L}_c$) in three respects: their language, axioms and rules of inference ($\mathcal{L}_c \subseteq \mathcal{L}_e, \mathcal{A}_c \subseteq \mathcal{A}_e, \mathcal{R}_c \subseteq \mathcal{R}_e$). These systems preserve all valid formula of the classical system, and therefore all previous valid formula remain valid as well ($\forall\varphi(\Sigma \models_c \varphi \Rightarrow \Sigma \models_e \varphi); \varphi \in \mathcal{L}_c$). So, for instance, modal logic extends classical system by the modal operators of necessity and possibility together with axioms and rules for them.

Deviations (*d*) are formal systems that share the language with the system of classical logic ($\mathcal{L}_c$) but that deviate in axioms and rules ($\mathcal{L}_c = \mathcal{L}_d, \mathcal{A}_c \neq \mathcal{A}_d, \mathcal{R}_c \neq \mathcal{R}_d$). Therefore, some formulae, which are valid in the classical system, are no longer valid in the deviant one ($\exists\varphi(\Sigma \models_c \varphi \wedge \Sigma \not\models_d \varphi); \varphi \in \mathcal{L}_c$). Such is the case of intuitionistic logic, in which the classical axiom $A \vee \neg A$ is no longer valid.

Inductive systems (*i*) are formal systems that share the language with the system of classical logic ($\mathcal{L}_c = \mathcal{L}_i$), but in which no formula which is valid by means of the inductive system is valid in the classical one ($\forall\varphi(\Sigma \models_i \varphi \Rightarrow \Sigma \not\models_c \varphi); \varphi \in \mathcal{L}_i$). Here the basis is the notion of 'inductive strength', and the idea is that 'an argument is inductively strong if its premisses give a certain degree of support, even if less than conclusive support, to its conclusion: if, that is, it is *improbable* that its premisses (Σ) should be true and its conclusion (φ) false' ($\textsc{not prob}(\Sigma \wedge \neg\varphi)$) [Haa78, page 17].

In each of these logical systems there is an underlying notion of logical consequence (or of derivability), which settles the validity of an argument within the system. While the first two categories pertain to formal systems which are deductive in nature, the third one concerns inductive ones. But still there may be several characterizations for both deductive and inductive kinds. For example, one deviant system, that of relevance logic renders the notion of classical consequence insufficient and asks for more: an argument in relevance logic must meet the requirement that the premisses be 'relevant' to its conclusion. As for inductive systems, another way of characterizing them is that for which 'it is improbable, *given that* the premisses (Σ) are true, that the conclusion is false ($\neg\varphi$)' [Haa78, page 17]. We may interpret this statement in terms of conditional probability as follows[2]: ($\text{NOT PROB}(\Sigma/\neg\varphi)$). Notice that deductive validity is a limiting case of inductive strength, where the probability of the premisses being true and the conclusion false is cero, for the first characterization, and where it is certain that the conclusion is true when the premisses are, for this second one.

In the overall, under this approach arguments may be assessed by deductive or inductive standards, and thus there may be deductively valid, inductively strong or neither. As we shall see later in this chapter (cf. section 5.1), this classification does not explicitly take into account abductive logic, but it can nevertheless be accommodated within.

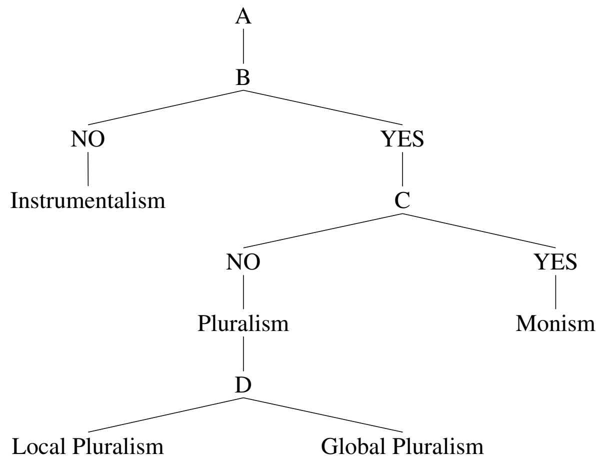
Positions: Instrumentalism, Monism, Pluralism

The position taken with respect to the demarcation of logic largely depends upon the answers given to metaphysical questions concerning the notion of correctness of a logical system, which in turn depend on the distinction between system-relative and extra-systematic validity/logical truth. Roughly speaking, a logical system is correct if the formal arguments (and formula) which are valid (logically true) in that system correspond to informal arguments (statements), which are valid (logically true) in the extra-systematic sense([Haa78, page 222]). Three positions are characterized by Haack, each of which is characterized by the answers (affirmative or negative) given to the following questions:

Questions:

- A: Does it make sense to speak of a logical system as correct or incorrect?
- B: Are there extra-systematic conceptions of validity/logical truth by means of which to characterize what is it for a logic to be correct?
- C: Is there one correct system?
- D: Must a logical system aspire to global application, i.e. to represent reasoning irrespective of subject-matter, or may a logic be locally correct. i.e. correct within a limited area of discourse?

[2] I thank the anonymous referee for this particular suggestion.



Thus, on the one hand, the instrumentalist position answers the first two questions negatively. It is based on the idea that the notion of 'correctness' for a system is inappropriate, and that one should rather be asking for its being more fruitful, useful, convenient... etc than another one. 'An instrumentalist will only allow the 'internal' question, whether a logical system is *sound*, whether, that is, all and only the theorems/syntactically valid arguments *of the system* are logically true/valid *in the system*'[Haa78, page 224]. On the other hand, both the monist and the pluralist answer these questions in the affirmative, the difference being that while the monist recognizes one and only one system of logic, the pluralist accepts a variety of them. Thus, they answer the third question opposite. Note that the distinction in these questions is only relevant for the classical logic vs. deviant logic dichotomy. The reason being that for a monist classical logic and its extensions are fragments of a 'correct system', and for a pluralist classical logic and its extensions are both 'correct'.

Likewise, while for an instrumentalist there are not extra-systematic conceptions of validity/logical truth by means of which to characterize what is to be a logic to be correct, for the monist as well as for the pluralist there are, either in the unitary fashion or in the pluralistic one. A further distinction made by the pluralist concerns the scope of application for a certain logical system. While a global pluralist endorses the view that a logical system must aspire to represent reasoning irrespective of subject-matter, a local pluralist supports the view that a logical system is only locally correct within a limited area of discourse.

The next question to analyze is the position taken by each of these stances with regard to whether deviant logics rival classical logic. In order to answer this question we have the following diagram:

(i) Formal argument which is	(ii) Valid in L
represents	corresponding to (iii)'s being
(iii) Informal argument	(iv) extra-systematically valid

On the one hand, the monist answers this question in the affirmative and supports the view that (i) aspires to represent (iii) in such a way that (ii) and (iv) do correspond in the 'correct logic'. On the other hand, the local pluralist answers this question in the negative by relativizing (iv) to specific areas of discourse and the global pluralist either fragments the relation between (i) and (iii) (that is, denies that the formal arguments of a deviant system represent the same informal arguments as those of classical logic) or fragments the relationship between (ii) and (iv) (denies that validity in the deviant logic is intended to correspond to extra-systematic validity as that to which validity in classical logic is intended to correspond). Finally, the instrumentalist rejects (iv) altogether.

Structural Logical Approach

This type of analysis (started in [Sco71]) was inspired in the works of logical consequence by Tarski [Tar83] and those of natural deduction by Gentzen [SD93, Pao02]. It describes a style of inference at a very abstract structural level, giving its pure combinatorics. It has proved very successful in artificial intelligence for studying different types of plausible reasoning ([KLM90]), and indeed as a general framework for non-monotonic consequence relations ([Gab85]). Another area where it has proved itself is dynamic semantics, where not one but many new notions of dynamic consequences are to be analyzed ([vBe96a]). The basic idea of logical structural analysis is the following:

> A notion of logical inference can be completely characterized by its basic combinatorial properties, expressed by structural rules.

Structural rules are instructions which tell us, e.g., that a valid inference remains valid when we insert additional premisses ('monotonicity'), or that we may safely chain valid inferences ('transitivity' or 'cut'). To understand this perspective in more detail, one must understand how it characterizes classical inference. In what follows we use logical sequents with a finite sequence of premisses to the left, and one conclusion to the right of the sequent arrow ($\Sigma \Rightarrow C$). While X, Y and Z are finite sets of formulae, A, B and C are single formula.

Classical Inference

The structural rules for classical inference are the following:

- **Reflexivity:** $C \Rightarrow C$
- **Contraction:**

$$\frac{X, A, Y, A, Z \Rightarrow C}{X, A, Y, Z \Rightarrow C}$$

- **Permutation:**

$$\frac{X, A, B, Y \Rightarrow C}{X, B, A, Y \Rightarrow C}$$

- **Monotonicity:**

$$\frac{X, Y \Rightarrow C}{X, A, Y \Rightarrow C}$$

- **Cut Rule:**

$$\frac{X, A, Y \Rightarrow C \quad Z \Rightarrow A}{X, Z, Y \Rightarrow C}$$

These rules state the following properties of classical consequence. Any premisse implies itself (reflexivity), deleting repeated premisses causes no trouble (contraction); premisses may be permuted without altering validity (permutation), adding new information does not invalidate previous conclusions (monotonicity), and premisses may be replaced by sequences of premisses implying them (cut). In all, these rules allow us to treat the premisses as a mere set of data without further relevant structure. This plays an important role in classical logic, witness what introductory textbooks have to say about "simple properties of the notion of consequence"[3]. Structural rules are also used extensively in completeness proofs[4].

These rules are structural in that they mention no specific symbols of the logical language. In particular, no connectives or quantifiers are involved. This makes the structural rules different from inference rules like, say, Conjunction of Consequents or Disjunction of Antecedents, which also fix the meaning of

[3]In [Men64, Page 30] the following simple properties of classical logic are introduced:

- If $\Gamma \subseteq \Delta$ and $\Gamma \vdash \phi$, then $\Delta \vdash \phi$.
- $\Gamma \vdash \phi$ iff there is a finite subset Δ of Γ such that $\Delta \vdash \phi$.
- If $\Gamma \vdash x_i$ (for all i) and $x_1, \ldots, x_n \vdash \phi$ then $\Gamma \vdash \phi$.

Notice that the first is a form of Monotonicity, and the third one of Cut.

[4]As noted in [Gro95, page46]: "In the Henkin construction for first-order logic, or propositional modal logic, the notion of maximal consistent set plays a major part, but it needs the classical structural rules. For example, Permutation, Contraction and Expansion enable you to think of the premisses of an argument as a set; Reflexivity is needed to show that for maximal consistent sets, membership and derivability coincide".

conjunction and disjunction. Under this approach, Haack's previous classification of extensions of logics is subsumed, for one rule may fit classical logic as well as extensions: propositional, first-order, modal, type-theoretic, etc. Each rule in the above list reflects a property of the set-theoretic definition of classical consequence ([Gro95]), which – with some abuse of notation – calls for inclusion of the intersection of the (models for the) premisses in the (models for the) conclusion:

$P_1, \ldots, P_n \Rightarrow C$ iff $P_1 \cap \ldots \cap P_n \subseteq C$.

Now, in order to prove that a set of structural rules *completely* characterizes a style of reasoning, representation theorems exist. For classical logic, one version was proved by van Benthem in [vBe91]:

PROPOSITION 1 *Monotonicity, Contraction, Reflexivity, and Cut completely determine the structural properties of classical consequence.*

Proof. Let R be any abstract relation between finite sequences of objects and single objects satisfying the classical structural rules. Now, define:

$$a^* = \{A \mid A \text{ is a finite sequence of objects such that ARa}\}.$$

Then, it is easy to show the following two assertions:

1 If $a_1, \ldots, a_k Rb$, then $a_1^* \cap \ldots \cap a_k^* \subseteq b^*$, using Cut and Contraction.

2 If $a_1^* \cap \ldots \cap a_k^* \subseteq b^*$, then $a_1, \ldots, a_k Rb$, using Reflexivity and Monotonicity. $\dashv$

Permutation is omitted in this theorem. And indeed, it turns out to be derivable from Monotonicity and Contraction.

We have thus shown that classical deductive inference, observes easy forms of reflexivity, contraction, permutation, monotonicity and cut. The representation theorem shows that these rules completely characterize this type of reasoning.

Non-Classical Inference

For non-classical consequences, classical structural rules may fail. Well-known examples are the ubiquitous 'non-monotonic logics'. However, this is not to say that no structural rules hold for them. The point is rather to find appropriate reformulations of classical principles (or even entirely new structural rules) that fit other styles of consequence. For example, many non-monotonic types of inference do satisfy a weaker form of monotonicity. Additions to the premisses are allowed only when these premisses imply them:

- Cautious Monotonicity:

$$\frac{X \Rightarrow C \qquad X \Rightarrow A}{X, A \Rightarrow C}$$

Dynamic inference is non-monotonic (inserting arbitrary new processes into a premisse sequence can disrupt earlier effects). But it also quarrels with other classical structural rules, such as Cut. But again, representation theorems exist. Thus, the 'update-to-test' dynamic style (once in which a process cannot be disrupted) and is characterized by the following restricted forms of monotonicity and cut, in which additions and omissions are licensed only to the left side:

- Left Monotonicity:

$$\frac{X \Rightarrow C}{A, X \Rightarrow C}$$

- Left Cut:

$$\frac{X \Rightarrow C \qquad X, C, Y \Rightarrow D}{X, Y \Rightarrow D}$$

For a broader survey and analysis of dynamic styles, see [Gro95, vBe96a]. For sophisticated representation theorems in the broader field of non-classical inference in artificial intelligence see [Mak93, KLM90]. Yet other uses of non-classical structural rules occur in relevance logic, linear logic, and categorial logics (cf. [DH93, vBe91]. [Gab94b]).

Characterizing a notion of inference in this way, determines its basic repertoire for handling arguments. Although this does not provide a more ambitious semantics, or even a full proof theory, it can at least provide valuable hints. The suggestive Gentzen style format of the structural rules turns into a sequent calculus, if appropriately extended with introduction rules for connectives. However, it is not always clear how to do so in a natural manner, as we will discuss later in connection with abduction.

The structural analysis of a logical inference is a metalevel explication which is based on structural rules and not on language, as it does not take into account logical connectives or constants, and in this respect differs from Haack's approach.

3. Abductive Explanatory Argument: A Logical Inference

Here are some preliminary remarks about the logical nature of abductive inference, which set the scene for our subsequent discussion. The standard textbook pattern of logical inference is this: *Conclusion C follows from a set of premisses P*. This format has its roots in the axiomatic tradition in mathematics that follows the deductive method, inherited from Euclid's *Elements*, in which from a certain set of basic axioms, all geometrical truths of elementary geometry are derived. This work is not only the first logical system of its kind, but it has been the model to follow in mathematics as well as in other formal scientific enterprises. Each proposition is linked, via proofs, to previous axioms, definitions and propositions. This method is forward chained, picturing a reasoning from a finite set of premisses to a conclusion, and it is completely certain and monotonic[5].

Moreover, there are at least two ways of thinking about validity in this setting, one semantic, based on the notions of model and interpretation (every model in which P is true makes C true), the other syntactic, based on a proof-theoretic derivation of C from P. Both explications suggest forward chaining from premisses to conclusions: $P \Rightarrow C$ and the conclusions generated are undefeasible. We briefly recall some features that make abduction a form of inference that does not fit easily into this format. All of them emerged in the course of our preceding chapter. Most prominently, in abduction, the conclusion is the given and the premisses (or part of them) are the output of the inferential process: $P \Leftarrow C$. Moreover, the *abduced* premisse has to be *consistent* with the background theory of the inference, as it has to be *explanatory*. And such explanations may undergo change as we modify our background theory. Finally, when different sets of premisses can be abduced as explanations, we need a notion of preference between them, allowing us to choose a *best* or *minimal* one. These various features, though non-standard when compared with classical logic, are familiar from neighbouring areas. For instance, there are links with classical accounts of explanation in the philosophy of science [Car55, Hem65], as well as recent research in artificial intelligence on various notions of common sense reasoning [McC80, Sho88, Gab96]. It has been claimed that this is an appropriate broader setting for general logic as well [vBe90], gaping back to the original program by Bernard Bolzano (1781–1848), in his "Wissenschaftslehre" [Bol73]. Indeed, our discussion of abduction in Peirce in the preceding chapter reflected a typical feature of pre-Fregean logic: boundaries between logic and general methodology were still rather fluid. In our view, current post-Fregean logical research is slowly moving back towards this same broader agenda. More concretely, we shall review the mentioned features of abduction in some further detail now, making a few strategic references to this broader literature.

Directions in Reasoning: Forward and Backward

Intuitively, a valid inference from, say, premisses P_1, P_2 to a conclusion C allows for various directions of thought. In a forward direction, given the premisses, we may want to draw some strongest, or rather, some most appropriate conclusion. (Notice incidentally, that the latter notion already introduces a certain dependence on context, and good sense: the strongest conclusion is simply

[5] This is not say however –as Hilbert would have liked to claim– that all mathematical reasoning may be reduced to axiomatics. As it is well know by the incompleteness results of Gödel, there are clear limitations to reasoning in mathematics through the axiomatic method. Moreover, the view of mathematics as an experimental, empirical science, found in philosophy[Lak76] as well as in recent work in computer science [Cha97], shows that axiomatics cannot exhaust all there is to mathematical reasoning.

$P_1 \wedge P_2$, but this will often be unsuited.) Classical logic also has a backward direction of thought, when engaged in refutation. If we know that C is false, then at least one of the premisses must be false. And if we know more, say the truth of P_1 and the falsity of the conclusion, we may even refute the specific premisse P_2. Thus, in classical logic, there is a duality between forward proof and backward refutation. This duality has been noted by many authors. It has even been exploited systematically by Beth when developing his refutation method of semantic tableaux [Bet69]. Read in one direction, closed tableaux are eventually failed analyses of possible counterexamples to an inference, read in another they can be arranged to generate a Gentzen-style sequent derivations of the inference (we shall be using tableaux in our next chapter, on computing abduction.) Beth's tableaux can be taken as a formal model of the historical opposition between methods of 'analysis' and 'synthesis' in the development of scientific argument (cf. chapter 1). Methodologically, the directions are different sides of the same coin, namely, some appropriate notion of inference.

Likewise, in abduction, we see an interplay of different directions. This time, though, the backward direction is not meant to provide refutations, but rather confirmations. We are looking for suitable premisses that would support the conclusion[6].

Our view of the matter is the following. In the final analysis, the distinction between directions is a relative one. What matters is not the direction of abduction, but rather an interplay of two things. As we have argued in chapter 2, one should distinguish between the choice of an underlying *notion of inference* $\Rightarrow$, and the independent issue as to the *search strategy* that we use over this. Forward reasoning is a bottom up use of $\Rightarrow$, while backward reasoning is a top-down use of $\Rightarrow$. In line with this, in this chapter, we shall concentrate on notions of inference $\Rightarrow$ leaving further search procedures to the next, more computational chapter 4. In this chapter the intuitively backward direction of abduction is not crucial to us, except as a pleasant manner of speaking. Instead, we concentrate on appropriate underlying notions of consequence for abduction.

Formats of Inference: Premisses and Background Theory

The standard format of logical inference is essentially binary, giving a transition from premisses to a conclusion:

$$\frac{P_1, \ldots, P_n}{C}$$

[6]In this case, a corresponding refutation would rather be a forward process: if the abduced premisse turns out false, it is discarded and an alternative hypothesis must be proposed. Interestingly, [Tij97] (a recent practical account of abduction in diagnostic reasoning) mixes both 'positive' confirmation of the observation to be explained with 'refutation' of alternatives.

These are 'local steps', which take place in the context of some, implicit or explicit, background theory (as we have seen in chapter 2). In this standard format, the background theory is either omitted, or lumped together with the other premisses. Often this is quite appropriate, especially when the background theory is understood. But sometimes, we do want to distinguish between different roles for different types of premisse, and then a richer format becomes appropriate. The latter have been proposed, not so much in classical logic, but in the philosophy of science, artificial intelligence, and informal studies on argumentation theory. These often make a distinction between explicit premisses and implicit *background assumptions*. More drastically, premisse sets, and even background theories themselves often have a hierarchical structure, which results in different 'access' for propositions in inference. This is a realistic picture, witness the work of cognitive psychologists like [Joh83].

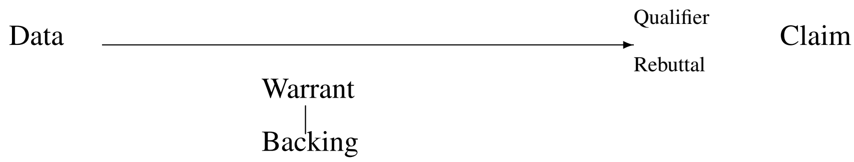
In Hempel's account of scientific explanation (cf. chapter 5) premisses play the role of either scientific laws, or of initial conditions, or of specific explanatory items, suggesting the following format:

$$\begin{array}{c}\text{Scientific laws + initial conditions + explanatory facts}\\ \Downarrow \\ \text{Observation}\end{array}$$

Further examples are found on the borderline of the philosophy of science and philosophical logic, in the study of conditionals. The famous 'Ramsey Test' presupposes revision of explicit beliefs in the background assumptions [Sos75, vBe94], which again have to be suitably structured. More elaborate hierarchical views of theories have been proposed in artificial intelligence and computer science. [Rya92] defines 'ordered theory presentations', which can be arbitrary rankings of principles involved in some reasoning practice. (Other implementations of similar ideas use labels for formulas, as in the labelled deductive systems of [Gab96].) While in Hempel's account, structuring the premisses makes sure that scientific explanation involves an interplay of laws and facts, Ryan's motivation is resolution of conflicts between premisses in reasoning, where some sentences are more resistant than others to revision. (This motivation is close to that of the Gärdenfors theory, to be discussed in chapter 8. A working version of these ideas is found in a study of abduction in diagnosis ([Tij97], which can be viewed as a version of our later account in this chapter with some preference structure added.) More structured views of premisses and theories can also be found in situation semantics, with its different types of 'constraints' that govern inference (cf. [PB83]).

In all these proposals, the theory over which inference takes place is not just a bag into which formulas are thrown indiscriminately, but an organized structure in which premisses have a place in a hierarchy, and play specific different roles. These additional features need to be captured in richer inferential formats for more complicated reasoning tasks. Intuitive 'validity' may be partly based on the type and status of the premisses that occur. We cite one more example, to elaborate what we have in mind.

In argumentation theory, an interesting proposal was made in [Tou58]. Toulmin's general notion of consequence was inspired on the kind of reasoning done by lawyers, whose claims need to be defended according to juridical procedures, which are richer than pure mathematical proof. Toulmin's format of reasoning contains the necessary tags for these procedures:



Every claim is defended from certain relevant data, by citing (if pressed) the background assumptions (one's 'warrant') that support this transition. (There is a dynamic process here. If the warrant itself is questioned, then one has to produce one's further 'backing'.) Moreover, indicating the purported strength of the inference is part of making any claim (whence the 'qualifier'), with a 'rebuttal' listing some main types of possible exception (rebuttal) that would invalidate the claim. [vBe94] relates this format to issues in artificial intelligence, as it seems to describe common sense reasoning rather well. Toulmin's model has also been proposed as a mechanism for intelligent systems performing explanation ([Ant89]).

Thus, once again, to model reasoning outside of mathematics, a richer format is needed. Notice that the above proposals are syntactic. It may be much harder to find purely semantic correlates to some of the above distinctions: as they seem to involve a reasoning procedure rather than propositional content. (For instance, even the distinction between individual facts and universal laws is not as straightforward as it might seem.) Various aspects of the Toulmin schema will return in what follows. For Toulmin, inferential strength is a parameter, to be set in accordance with the subject matter under discussion. (Interestingly, content-dependence of reasoning is also a recurrent finding of cognitive psychologists: cf. the earlier-mentioned [Joh83].) In chapter 2, we have already defended exactly the same strategy for abduction. Moreover, the procedural flavor of the Toulmin schema fits well with our product-process distinction.

As for the basic building blocks of abductive explanatory inference, in the remainder of this book, we will confine ourselves to a ternary format:

$$\Theta \mid \alpha \Rightarrow \varphi$$

This modest step already enables us to demonstrate a number of interesting departures from standard logical systems. Let us recall some considerations from chapter 2 motivating this move. The theory Θ needs to be explicit for a number of reasons. Validity of an abductive inference is closely related to the background theory, as the presence of some other explanation β in Θ may actually disqualify α as an explanation. Moreover, what we called 'triggers' of explanation are specific conditions on a theory Θ and an observation φ. A fact may need explanation with respect to one theory, but not with respect to another. Making a distinction between Θ and α allows us to highlight the specific explanation (which we did not have before), and control different forms of explanation (facts, rules, or even new theories). But certainly, our accounts would become yet more sensitive if we worked with some of the above richer formats.

Inferential Strength: A Parameter

At first glance, once we have Tarski's notion of truth, logical consequence seems an obvious defined notion. A conclusion follows if it is true in all models where the premisses are true. But the contemporary philosophical and computational traditions have shown that natural notions of inference may need more than truth in the above sense, or may even hinge on different properties altogether. For example, among the candidates that revolve around truth, statistical inference requires not total inclusion of premisse models in conclusion models, but only a significant overlap, resulting in a high degree of certainty. Other approaches introduce new semantic primitives. Notably, Shoham's notion of causal and default reasoning ([Sho88]) introduces a preference order on models, requiring only that the *most preferred models* of Σ be included in the models of φ.

More radically, dynamic semantics replaces the notion of truth by that of *information change*, aiming to model the flow of information. This move leads to a redesign for Tarski semantics, with e.g. quantifiers becoming actions on assignments ([vBC94]). This logical paradigm has room for many different inferential notions ([Gro95, vBe96a]). An example is the earlier mentioned update-to-test-consequence:

"process the successive premisses in Σ, thereby absorbing their informational content into the initial information state. At the end, check if the resulting state is rich enough to satisfy the conclusion φ".

Informational content rather than truth is also the key semantic property in situation theory ([PB83]). In addition to truth-based and information-based approaches, there are, of course, also various proof-theoretic variations on standard consequence. Examples are default reasoning: "φ is provable unless and until φ is disproved" ([Rei80]), and indeed Hempel's hypothetico-deductive model of scientific inference itself.

All these alternatives agree with our analysis of abductive explanatory inference. On our view, abduction is not a new notion of inference. It is rather a topic-dependent practice of scientific reasoning, which can be supported by various members of the above family. In fact, it is appealing to think of abductive inference in several respects, as inference involving preservation of both truth and *explanatory power*. In fact, appropriately defined, both might turn out equivalent. It has also been argued that since abduction is a form of reversed deduction, just as deduction is truth-preserving, abduction must be falsity-preserving ([Mic94]). However, [Fla95] gives convincing arguments against this particular move. Moreover, as we have already discussed intuitively, abduction is not just deduction in reverse.

Our choice here is to study abductive inference in more depth as a strengthened form of classical inference. This is relevant, it offers nice connections with artificial intelligence and the philosophy of science, and it gives a useful simple start for a broader systematic study of abductive inference. One can place this choice in a historical context, namely the work of Bernard Bolzano, a nineteenth century philosopher and mathematician (and theologian) engaged in the study of different varieties of inference. We provide a brief excursion, providing some perspective for our later technical considerations.

Bolzanos's Program

Bolzano's notion of deducibility (*Ableitbarkeit*) has long been recognized as a predecessor of Tarski's notion of logical consequence ([Cor75]). However, the two differ in several respects, and in our broader view of logic, they even appear radically different. These differences have been studied both from a philosophical ([Tho81]) and from a logical point of view ([vBe84a]).

One of Bolzano's goals in his theory of science ([Bol73]), was to show why the claims of science form a theory as opposed to an arbitrary set of propositions. For this purpose, he defines his notion of deducibility as a logical relationship extracting conclusions from premisses forming *compatible propositions*, those for which some set of ideas make all propositions true when uniformly substituted throughout. In addition, compatible propositions must share *common ideas*. Bolzano's use of 'substitutions' is of interest by itself, but for our purposes here, we will identify these (somewhat roughly) with the standard use of 'models'. Thompson attributes the difference between Bolzano's consequence and Tarski's to the fact that the former notion is epistemic while the latter is ontological. These differences have strong technical effects. With Bolzano, the premisses must be consistent (sharing at least one model), with Tarski, they need not. Therefore, from a contradiction, everything follows for Tarski, and nothing for Bolzano.

Restated for our ternary format, then, Bolzano's notion of deducibility reads as follows (cf. [vBe84a]):

$\Theta \mid \alpha \Rightarrow \varphi$ if
(1) The conjunction of Θ and α is consistent.
(2) Every model for Θ plus α verifies φ.

Therefore, Bolzano's notion may be seen (anachronistically) as Tarski's consequence plus the additional condition of consistency. Bolzano does not stop here. A finer grain to deducibility occurs in his notion of *exact deducibility*, which imposes greater requirements of 'relevance'. A modern version, involving inclusion-minimality for sets of abducibles, may be transcribed (again, with some historical injustice) as:

$\Theta \mid \alpha \Rightarrow^{+} \varphi$ if
(1) $\Theta \mid \alpha \Rightarrow \varphi$
(2) There is no proper subset of α, α', such that $\Theta \mid \alpha' \Rightarrow \varphi$.

That is, in addition to consistency with the background theory, the premisse set α must be 'fully explanatory' in that no subpart of it would do the derivation. Notice that this leads to non-monotonicity. Here is an example:

$\Theta \mid a \rightarrow b, a \Rightarrow^{+} b$
$\Theta \mid a \rightarrow b, a, b \rightarrow c \not\Rightarrow^{+} b$

Bolzano's agenda for logic is relevant to our study of abductive reasoning (and the study of general non-monotonic consequence relations) for several reasons. It suggests the methodological point that what we need is not so much proliferation of different logics as a better grasp of different styles of consequence. Moreover, his work reinforces an earlier claim, that truth is not all there is to understanding explanatory reasoning. More specifically, his notions still have interest. For example, exact deducibility has striking similarities to explanation in philosophy of science (cf. chapter 5).

Abductive Explanatory Inference as Deduction in Reverse

In this section we define abductive explanatory inference as a strengthened form of classical inference. Our proposal will be in line with abduction in artificial intelligence, as well as with the Hempelian account of explanation. We will motivate our requirements with our very simple rain example, presented here in classical propositional logic:

$\Theta : r \rightarrow w, s \rightarrow w$
$\varphi : w$

The first condition for a formula α to count as an explanation for φ with respect to Θ is the inference requirement. Many formulas would satisfy this condition. In addition to earlier-mentioned obvious explanations (r: rain, s: sprinklers-on), one might take their conjunction with any other formula, even if the latter is inconsistent with Θ (e.g. $r \wedge \neg w$). One can take the fact itself (w), or, one can introduce entirely new facts and rules (say, there are children playing with water, and this causes the lawn to get wet).

Inference: $\Theta, \alpha \models \varphi$

α's: r, s, $r \wedge s$, $r \wedge z$, $r \wedge \neg w$, $s \wedge \neg w$, w, $[c, c \rightarrow w]$, $\Theta \rightarrow w$.

Some of these 'abductive explanations' must be ruled out from the start. We therefore impose a consistency requirement on the left hand side, leaving only the following as possible abductive explanations:

Consistency: Θ, α is consistent.

α's: r, s, $r \wedge s$, $r \wedge z$, w, $[c, c \rightarrow w]$, $\Theta \rightarrow w$.

An abductive explanation α is only *necessary*, if φ is not already entailed by Θ. Otherwise, any consistent formula will count as an abductive explanation. Thus we repeat an earlier trigger for abduction: $\Theta \not\models \varphi$. By itself, this does not rule out any potential abducibles on the above list (as it does not involve the argument α.) But also, in order to avoid what we may call *external explanations* –those that do not use the background theory at all (like the explanation involving children in our example) –, it must be required that α be insufficient for explaining φ by itself ($\alpha \not\models \varphi$). In particular this condition avoids the trivial reflexive explanation $\varphi \not\Rightarrow \varphi$. Then only the following explanations are left in our list of examples:

Explanation $\Theta \not\models \varphi$, $\alpha \not\models \varphi$

α's: r, s, $r \wedge s$, $r \wedge z$, $\Theta \rightarrow w$.

Now both Θ and α contribute to explaining φ. However, we are still left with some formulas that do not seem to be genuine explanations ($r \wedge z, \Theta \rightarrow w$). Therefore, we explore a more sensitive criterion, admitting only 'the best explanation'.

Selecting the Best Explanation

Intuitively, a reasonable ground for choosing a statement as the best explanation, is its simplicity. It should be minimal, i.e. as weak as possible in performing its job. This would lead us to prefer r over $r \wedge z$ in the preceding example. As Peirce puts it, we want the explanation that *"adds least to what has been observed"* (cf. [CP, 6.479]). The criterion of simplicity has been extensively considered both in the philosophy of science and in artificial intelligence. But its precise formulation remains controversial, as measuring simplicity can be a tricky matter. One attempt to capture simplicity in a logical way is as follows:

Weakest Abductive Explanation:

α is the weakest abductive explanation for φ with respect to Θ iff

(i) $\Theta, \alpha \models \varphi$

(ii) For all other formulas β such that $\Theta, \beta \models \varphi$, $\models \beta \rightarrow \alpha$.

This definition makes the explanations r and s almost the weakest in the above example, just as we want. Almost, but not quite. For, the explanation $\Theta \rightarrow w$, a trivial solution, turns out to be the minimal one. The following is a folklore observation to this effect:

FACT 1 *Given any theory Θ and observation φ to be explained from it, $\alpha = \Theta \rightarrow \varphi$ is the weakest abductive explanation.*

Proof. Obviously, we have (i) $\Theta, \Theta \rightarrow \varphi \models \varphi$. Moreover, let α' be any other explanation. This says that $\Theta, \alpha' \models \varphi$. But then we also have (by conditionalizing) that $\alpha' \models \Theta \rightarrow \varphi$, and hence $\models \alpha' \rightarrow (\Theta \rightarrow \varphi)$ $\dashv$

That $\Theta \rightarrow \varphi$ is a solution that will always count as an explanation in a deductive format was noticed by several philosophers of science ([Car55]). It has been used as an argument to show how the issue would impose restrictions on the syntactic form of abducibles. Surely, in this case, the explanation seems too complex to count. We will therefore reject this proposal, noting also that it fails to recognize (let alone compare) intuitively 'minimal' explanations like r and s in our running example.

Other criteria of minimality exist in the literature. One of them is based on preference orderings. The best explanation is the most preferred one, given an explicit ordering of available assertions. In our example, we could define an order in which inconsistent explanations are the least preferred, and the simplest the most. These preference approaches are quite flexible, and can accommodate various working intuitions. However, they may still depend on many factors, including the background theory. This seems to fall outside a logical framework, referring rather to further 'economic' decision criteria like utilities. A case in point is Peirce's 'economy of research' in selecting a most promising hypothesis. What makes a hypothesis good or best has no easy answer. One may appeal to criteria of simplicity, likelihood, or predictive power. To complicate matters even further, we often do not compare (locally) quality of explanations given a fixed theory, but rather (globally) whole packages of 'theory + explanation'. This perspective gives a much greater space of options. As we have not been able to shed a new light from logic upon these matters, we will ignore these dimensions here.

Further study would require more refined views of theory structure and reasoning practice, in line with some of the earlier references[7], or even more ambitiously, following current approaches to 'verisimilitude' in the philosophy of science (cf. [Kui87]).

[7] Preferences over models (though not over statements) will be mentioned briefly as providing one possible inference mechanism for abduction.

We conclude with one final observation. perhaps one reason why the notion of 'minimality' has proved so elusive is again our earlier product-process distinction. Philosophers have tried to define minimality in terms of intrinsic properties of statements and inferences as products. But it may rather be a process-feature, having to do with computational effort in some particular procedure performing abduction. Thus, one and the same statement might be minimal in one abduction, and non-minimal in another.

Abductive Explanatory Characterization Styles

Following our presentation of various requirements for an abductive explanation, we make things more concrete for further reference. We consider five versions of abductive explanations making up the following styles: plain, consistent, explanatory, minimal and preferential, defined as follows:

Abductive Explanatory Styles

Given Θ (a set of formulae) and φ (a sentence), α is an *abductive explanation* if:

Plain :

(i) $\Theta, \alpha \models \varphi$.

Consistent :

(i) $\Theta, \alpha \models \varphi$,

(ii) Θ, α consistent.

Explanatory :

(i) $\Theta, \alpha \models \varphi$,

(ii) $\Theta \not\models \varphi$,

(iii) $\alpha \not\models \varphi$.

Minimal :

(i) $\Theta, \alpha \models \varphi$,

(ii) α is the weakest such abductive explanation (not equal to $\Theta \rightarrow \varphi$).

Preferential :

(i) $\Theta, \alpha \models \varphi$,

(ii) α is the best abductive explanation according to some given preferential ordering.

We can form other combinations, of course, but these will already exhibit many characteristic phenomena. Note that these requirements do not depend on classical consequence. For instance, in Chapter 5, the consistency and the explanatory requirements work just as well for statistical inference. The former then also concerns the explanandum φ. (For, in probabilistic reasoning it is possible to infer two contradictory conclusions even when the premisses are consistent.) The latter helps capture when an explanation helps raise the probability of the explanandum.

A full version of abduction would make the formula to be abduced part of the derivation, consistent, explanatory, and the best possible one. However, instead of incorporating all these conditions at once, we shall consider them one by one. Doing so clarifies the kind of restriction each requirement adds to the notion of plain abduction. Our standard versions will base these requirements on classical consequence underneath. But we also look briefly toward the end at versions involving other notions of consequence. We will find that our various notions of abduction have advantages, but also drawbacks, such as an increase of complexity for explanatory reasoning as compared with classical inference.

Up to now, we can say that abductive inference may be characterized by reversed deduction plus additional conditions. However, is this all we can say about the logic of abduction? This definition does not really capture the rationality principles behind this type of reasoning, like its non-monotonic feature we have talked about.

In what follows, our aim is to present the characterization of abductive inference from the structural perspective we introduced earlier in this chapter. This approach has become popular across a range of non-standard logics. Our systematic analysis will explore different abductive styles from this perspective.

4. Abductive Explanatory Inference: Structural Characterization

Consistent Abductive Explanatory Inference

We recall the definition:

$\Theta \mid \alpha \Rightarrow \varphi$ iff
(i) $\Theta, \alpha \models \varphi$
(ii) Θ, α are consistent

The first thing to notice is that the two items to the left behave *symmetrically*:

$\Theta \mid \alpha \Rightarrow \varphi \qquad \text{iff} \qquad \alpha \mid \Theta \Rightarrow \varphi$

Indeed, in this case, we may technically simplify matters to a binary format after all: $X \Rightarrow C$, in which X stands for the conjunction of Θ and α, and C for φ. To bring these in line with the earlier-mentioned structural analysis of nonclassical logics, we view X as a finite sequence $X_1 \ldots, X_k$ of formulas and C as a single conclusion.

Classical Structural Rules

Of the structural rules for classical consequence, contraction and permutation hold for consistent abduction. But reflexivity, monotonicity and cut fail, witness by the following counterexamples:

- Reflexivity: $p \wedge \neg p \not\Rightarrow p \wedge \neg p$
- Monotonicity: $p \Rightarrow p$, but $p, \neg p \not\Rightarrow p$
- Cut: $p, \neg q \Rightarrow p$, and $p, q \Rightarrow q$, but $p, \neg q, q \not\Rightarrow q$

New Structural Rules

Here are some restricted versions of the above failed rules, and some others that are valid for consistent abduction:

1 Conditional Reflexivity (**CR**)

$$\frac{X \Rightarrow B}{X \Rightarrow X_i} \qquad 1 \leq i \leq k$$

2 Simultaneous Cut (**SC**)

$$\frac{U \Rightarrow A_1 \ldots U \Rightarrow A_k \qquad A_1, \ldots, A_k \Rightarrow B}{U \Rightarrow B}$$

3 Conclusion Consistency (**CC**)

$$\frac{U \Rightarrow A_1 \ldots U \Rightarrow A_k}{A_1, \ldots, A_k \Rightarrow A_i} \qquad 1 \leq i \leq k$$

These rules state the following. Conditional Reflexivity requires that the sequence X derive something else ($X \Rightarrow B$), as this ensures consistency. Simultaneous Cut is a combination of Cut and Contraction in which the sequent $A_1, \ldots, A_k$ may be omitted in the conclusion when each of its elements A_i is consistently derived by U and this one in its turn consistently derives B. Conclusion Consistency says that a sequent $A_1, \ldots, A_k$ implies its elements if each of these are implied consistently by something (U arbitrary), which is another form of reflexivity.

PROPOSITION 2 *These rules are sound for consistent abduction.*

Proof. In each of these three cases, it is easy to check by simple set-theoretic reasoning that the corresponding classical consequence holds. Therefore, the only thing to be checked is that the premisses mentioned in the conclusions of these rules must be consistent. For Conditional Reflexivity, this is because X already consistently implied something. For Simultaneous Cut, this is because U already consistently implied something. Finally, for Conclusion Consistency, the reason is that U must be consistent, and it is contained in the intersection of all the A_i, which is therefore consistent, too. $\dashv$

A Representation Theorem

The given structural rules in fact characterize consistent abduction:

PROPOSITION 3 *A consequence relation satisfies structural rules 1 (CR), 2 (SC), 3 (CC) iff it is representable in the form of consistent abduction.*

Proof. Soundness of the rules was proved above. Now consider the completeness direction. Let $\Rightarrow$ be any abstract relation satisfying 1, 2, 3. Define for any proposition A,

$$A^* = \{X \mid X \Rightarrow A\}$$

We now show the following statement of adequacy for this representation:

Claim. $A_1, \ldots, A_k \Rightarrow B$ iff $\emptyset \subset A_1^* \cap \ldots \cap A_k^* \subseteq B^*$.

Proof. 'Only if'. Since $A_1, \ldots, A_k \Rightarrow B$, by Rule 1 (CR) we have $A_1, \ldots, A_k \Rightarrow A_i$ $(1 \leq i \leq k)$. Therefore, $A_1, \ldots, A_k \in A_i^*$, for each i with $1 \leq i \leq k$, which gives the proper inclusion. Next, let U be any sequence in the intersection of all A_i^*, for $1, \ldots, k$. That is, $U \Rightarrow A_1, \ldots, U \Rightarrow A_k$. By Rule 2 (SC), $U \Rightarrow B$, i.e. $U \in B^*$, and we have shown the second inclusion.

'If'. Using the assumption of non-emptiness, let, say, $U \in \bigcap A_i^*$, for $1, \ldots, k$. i.e. $U \Rightarrow A_1, \ldots, U \Rightarrow A_k$. By Rule 3 (CC), $A_1, \ldots, A_k \Rightarrow A_i$ $(1 \leq i \leq k)$. By the second inclusion then, $A_1, \ldots, A_k \in B^*$. By the definition of the function *, this means that $A_1, \ldots, A_k \Rightarrow B$. ⊣

More Familiar Structural Rules

The above principles characterize consistent abduction. Even so, there are more familiar structural rules that are valid as well, including modified forms of Monotonicity and Cut. For instance, it is easy to see that $\Rightarrow$ satisfies a form of modified monotonicity: B may be added as a premisse if this addition does not endanger consistency. And the latter may be shown by their 'implying' any conclusion:

- Modified Monotonicity:

$$\frac{X \Rightarrow A \qquad X, B \Rightarrow C}{X, B \Rightarrow A}$$

As this was not part of the above list, we expect some derivation from the above principles. And indeed there exists one:

- Modified Monotonicity Derivation:

$$\dfrac{\dfrac{X, B \Rightarrow C}{X, B \Rightarrow X_i' s}\,1 \qquad X \Rightarrow A}{X, B \Rightarrow A}\,2$$

These derivations also help in seeing how one can reason perfectly well with non-classical structural rules. Another example is the following valid form of Modified Cut:

- Modified Cut

$$\frac{X \Rightarrow A \qquad U, A, V \Rightarrow B \qquad U, X, V \Rightarrow C}{U, X, V \Rightarrow B}$$

This may be derived as follows:

- Modified Cut Derivation

$$\dfrac{\dfrac{U,X,V \Rightarrow C}{U,X,V \Rightarrow U's,V's}\,1 \qquad \dfrac{\dfrac{\dfrac{U,X,V \Rightarrow C}{U,X,V \Rightarrow X_i's}\,1 \qquad X \Rightarrow A}{U,X,V \Rightarrow A}\,2 \qquad U,A,V \Rightarrow B}{U,X,V \Rightarrow B}\,2$$

Finally, we check some classically structural rules that do remain valid as they stand, showing the power of Rule (3):

- Permutation

$$\dfrac{\dfrac{\dfrac{X,A,B,Y \Rightarrow C}{X,A,B,Y \Rightarrow X \quad X,A,B,Y \Rightarrow A \quad X,A,B,Y \Rightarrow B \quad X,A,B,Y \Rightarrow Y}\,1}{X,B,A,Y \Rightarrow X \quad X,B,A,Y \Rightarrow B \quad X,B,A,Y \Rightarrow A \quad X,B,A,Y \Rightarrow Y}\,3 \qquad X,A,B,Y \Rightarrow C}{X,B,A,Y \Rightarrow C}\,2$$

- Contraction (one sample case)

$$\dfrac{\dfrac{\dfrac{X,A,A,Y \Rightarrow B}{X,A,A,Y \Rightarrow X_i's,A,Y_i's}\,1}{X,A,Y \Rightarrow X_i's,A,Y_i's}\,3 \qquad X,A,A,Y \Rightarrow B}{X,A,Y \Rightarrow B}\,2$$

Thus, consistent abductive inference defined as classical consequence plus the consistency requirement has appropriate forms of reflexivity, monotonicity, and cut for which it is assured that the premisses remain consistent. Permutation and contraction are not affected by the consistency requirement, therefore the classical forms remain valid. More generally, the preceding examples show simple ways of modifying all classical structural principles by putting in one extra premisse ensuring consistency.

Simple as it is, our characterization of this notion of inference does provide a complete structural description of Bolzano's notion of deducibility introduced earlier in this chapter (section 3.3).

Explanatory Abductive (Explanatory) Inference

Explanatory abductive explanatory inference (explanatory abduction, for short)was defined as plain abduction ($\Theta, \alpha \models \varphi$) plus two conditions of necessity ($\Theta \not\models \varphi$) and insufficiency ($\alpha \not\models \varphi$). However, we will consider a weaker version (which only considers the former condition) and analyze its structural rules. This is actually somewhat easier from a technical viewpoint. The full version remains of general interest though, as it describes the 'necessary collaboration' of two premisses set to achieve a conclusion. It will be analyzed further in chapter 5 in connection with philosophical models of scientific explanation. We rephrase our notion as:

Weak Explanatory Abduction:
$\Theta \mid \alpha \Rightarrow \varphi$ iff
(i) $\Theta, \alpha \models \varphi$
(ii) $\Theta \not\models \varphi$

The first thing to notice is that we must leave the binary format of premisses and conclusion. This notion is non-symmetric, as Θ and α have different roles. Given such a ternary format, we need a more finely grained view of structural rules. For instance, there are now two kinds of monotonicity, one when a formula is added to the explanations and the other one when it is added to the theory:

- Monotonicity for Abductive Explanations:

$$\frac{\Theta \mid \alpha \Rightarrow \varphi}{\Theta \mid \alpha, A \Rightarrow \varphi}$$

- Monotonicity for Theories:

$$\frac{\Theta \mid \alpha \Rightarrow \varphi}{\Theta, A \mid \alpha \Rightarrow \varphi}$$

The former is valid, but the latter is not. (A counterexample is: $p \mid q, r \Rightarrow q$ but $p, q \mid q, r \not\Rightarrow q$). Monotonicity for explanations states that an explanation for a fact does not get invalidated when we strengthen it, as long as the theory is not modified.

Here are some valid principles for weak explanatory abduction.

- Weak Explanatory Reflexivity

$$\frac{\Theta \mid \alpha \Rightarrow \varphi}{\Theta \mid \varphi \Rightarrow \varphi}$$

- Weak Explanatory Cut

$$\frac{\Theta \mid \alpha, \beta \Rightarrow \varphi \qquad \Theta \mid \alpha \Rightarrow \beta}{\Theta \mid \alpha \Rightarrow \varphi}$$

In addition, the classical forms of contraction and permutation are valid on each side of the bar. Of course, one should not permute elements of the theory with those in the explanation slot, or vice versa. We conjecture that the given principles completely characterize the weak explanatory abduction notion, when used together with the above valid form of monotonicity.

Structural Rules with Connectives

Pure structural rules involve no logical connectives. Nevertheless, there are natural connectives that may be used in the setting of abductive consequence. For instance, all Boolean operations can be used in their standard meaning. These, too, will give rise to valid principles of inference. In particular, the following well-known classical laws hold for all notions of abductive inference studied so far:

- Disjunction of Θ-antecedents:

$$\frac{\Theta_1 \mid A \Rightarrow \varphi \qquad \Theta_2 \mid A \Rightarrow \varphi}{\Theta_1 \vee \Theta_2 \mid A \Rightarrow \varphi}$$

- Conjunction of Consequents

$$\frac{\Theta \mid A \Rightarrow \varphi_1 \qquad \Theta \mid A \Rightarrow \varphi_2}{\Theta \mid A \Rightarrow \varphi_1 \wedge \varphi_2}$$

These rules will play a role in our proposed calculus for abduction, as we will show later on.

Another way of expressing monotonicity with the aid of negation and classical derivability is as follows:

- Monotonicity:

$$\frac{\Theta \mid \alpha \Rightarrow \varphi \qquad \Theta \mid \alpha \not\vdash \neg\beta}{\Theta \mid \alpha, \beta \Rightarrow \varphi}$$

We conclude a few brief points on the other versions of abduction on our list. We have not undertaken to characterize these in any technical sense.

Minimal and Preferential Abductive Explanatory Inference

Consider our versions of 'minimal' abduction. One said that $\Theta, \alpha \models \varphi$ and α is the weakest such explanation. By contrast, preferential abduction said that $\Theta, \alpha \models \varphi$ and α is the best explanation according to some given preferential ordering. For the former, with the exception of the above disjunction rule for antecedents, no other rule that we have seen is valid. But it does satisfy the following form of transitivity:

- Transitivity for Minimal Abduction:

$$\frac{\Theta \mid \alpha \Rightarrow \varphi \qquad \Theta \mid \beta \Rightarrow \alpha}{\Theta \mid \beta \Rightarrow \varphi}$$

For preferential abduction, on the other hand, no structural rule formulated so far is valid. The reason is that the relevant preference order amongst formulas, in itself needs to be captured in the formulation of our inference rules. A valid formulation of monotonicity would then be something along the following lines:

- Monotonicity for Preferential Abduction:

$$\frac{\Theta \mid \alpha \Rightarrow \varphi \qquad \alpha, \beta < \alpha}{\Theta \mid \alpha, \beta \Rightarrow \varphi}$$

In our opinion, this is no longer a structural rule, since it adds a mathematical relation ($<$ for a preferential order) that cannot in general be expressed in terms of the consequence itself. This is a point of debate, however, and its solution depends on what each logic artisan is willing to represent in a logic. In any case, this format is beyond what we will study in this book.

Structural Rules for Nonstandard Inference

All abductive versions so far had classical consequence underneath. In this section, we briefly explore structural behaviour when the underlying notion of inference is non standard, as in preferential entailment. Moreover, we throw in some words about structural rules for abduction in logic programming, and for induction.

Preferential Reasoning

Interpreting the inferential parameter as preferential entailment means that $\Theta, \alpha \Rightarrow \varphi$ if (only) the most preferred models of $\Theta \cup \alpha$ are included in the models of φ. This leads to a completely different set of structural rules. Here are some valid examples, transcribed into our ternary format from [KLM90]:

- Reflexivity: $\Theta, \alpha \Rightarrow \alpha$

- Cautious Monotonicity:

$$\frac{\Theta \mid \alpha \Rightarrow \beta \qquad \Theta \mid \alpha \Rightarrow \gamma}{\Theta \mid \alpha, \beta \Rightarrow \gamma}$$

- Cut:

$$\frac{\Theta \mid \alpha, \beta \Rightarrow \gamma \qquad \alpha \Rightarrow \beta}{\Theta \mid \alpha \Rightarrow \gamma}$$

- Disjunction:

$$\frac{\Theta \mid \alpha \Rightarrow \varphi \qquad \Theta \mid \beta \Rightarrow \varphi}{\Theta \mid \alpha \vee \beta \Rightarrow \varphi}$$

It is interesting to see in greater detail what happens to these rules when we add our further conditions of 'consistency' and 'explanation'. In all, what happens is merely that we get structural modifications similar to those found earlier on for classical consequence. Thus, a choice for a preferential proof engine, rather than classical consequence, seems orthogonal to the behavior of abduction.

Structural rules for Prolog Computation

An analysis via structural rules may be also performed for notions of $\Rightarrow$ with a more procedural flavor. In particular, the earlier-mentioned case of Prolog computation obeys clear structural rules (cf. [vBe92, Kal95, Min90]). Their format is somewhat different from classical ones, as one needs to represent more of the Prolog program structure for premisses, including information on rule heads. (Also, Kalsbeek [Kal95] gives a complete calculus of structural rules for logic programming including such control devices as the cut operator !). The characteristic expressions of a Gentzen style sequent calculus for these systems (in the reference above) are sequents of the form $[P] \Rightarrow \varphi$, where P is a (propositional, Horn clause) program and φ is an atom. A failure of a goal is expressed as $[P] \Rightarrow \neg\varphi$ (meaning that φ finitely fails). In this case, valid monotonicity rules must take account of the place in which premisses are added, as Prolog is sensitive to the order of its program clauses. Thus, of the following rules, the first one is valid, but the second one is not:

- Right Monotonicity

$$\frac{[P] \Rightarrow \varphi}{[P;\beta] \Rightarrow \varphi}$$

- Left Monotonicity

$$\frac{[P] \Rightarrow \varphi}{[\beta;P] \Rightarrow \varphi}$$

  Counterexample: $\beta = \varphi \leftarrow \varphi$

The question of complete structural calculi for abductive logic programming will not be addressed in this book, we will just mention that a natural rule for an 'abductive update' is as follows:

- Atomic Abductive Update

$$\frac{[P] \Rightarrow \neg\varphi}{[P;\varphi] \Rightarrow \varphi}$$

We will briefly return to structural rules for abduction as a process in the next chapter.

Structural Rules For Induction

Unlike abduction, enumerative induction is a type of inference that explains a set of observations, and makes a prediction for further ones (cf. our discussion in chapter 2). Our previous rule for conjunction of consequents already suggests how to give an account for further observations, provided that we interpret the commas below as conjunction amongst formulae (in the usual Gentzen calculus, commas to the right are interpreted rather as disjunctions):

$$\frac{\alpha \Rightarrow \varphi_1 \qquad \alpha \Rightarrow \varphi_2}{\alpha \Rightarrow \varphi_1, \varphi_2}$$

That is, an inductive explanation α for φ_1 remains an explanation when a formula φ_2 is added, provided that α also accounts for it separately. Note that this rule is a kind of monotonicity, but this time the increase is on the conclusion set rather than on the premisse set. More generally, an inductive explanation α for a set of formulae remains valid for more input data ψ when it explains it:

- (Inductive) Monotonicity on Observations

$$\frac{\Theta \mid \alpha \Rightarrow \varphi_1, \ldots, \varphi_n \qquad \Theta \mid \alpha \Rightarrow \psi}{\Theta \mid \alpha \Rightarrow \varphi_1, \ldots, \varphi_n, \psi}$$

In order to put forward a set of rules characterizing inductive explanation, a further analysis of its properties should be made, and this falls beyond the scope of this thesis. What we anticipate however, is that a study of enumerative induction from a structural point of view will bring yet another twist to the standard structural analysis, that of giving an account of changes in conclusions.

Further Logical Issues

Our analysis so far has only scratched the surface of a broader field. In this section we discuss a number of more technical logical aspects of abductive styles of inference. This identifies further issues that seem relevant to understanding the logical properties of abduction.

Completeness

The usual completeness theorems have the following form:

$\Theta \models \varphi \qquad$ iff $\qquad \Theta \vdash \varphi$

With our ternary format, we would expect some similar equivalence, with a possibly different treatment of premisses on different sides of the comma:

$\Theta, \alpha \models \varphi \qquad$ iff $\qquad \Theta, \alpha \vdash \varphi$

Can we get such completeness results for any of the abductive versions we have described so far? Here are two extremes.

The representation arguments for the above characterizations of abduction may be reworked into completeness theorems of a very simple kind. (This works just as in [vBe96a], chapter 7). In particular, for consistent abduction, our earlier argument essentially shows that $\Theta, \alpha \Rightarrow \varphi$ follows from a set of ternary sequents Φ iff it can be derived from Φ using only the derivation rules (CR), (SC), (CC) above.

These representation arguments may be viewed as 'poor man's completeness proofs', for a language without logical operators. Richer languages arise by adding operators, and completeness arguments need corresponding 'upgrading' of the representations used. (Cf. [Kur95] for an elaborate analysis of this upward route for the case of categorial and relevance logics. [Gro95] considers the same issue in detail for dynamic styles of inference.) At some level, no more completeness theorems are to be expected. The complexity of the desired proof theoretical notion $\vdash$ will usually be recursively enumerable (Σ_1^0). But, our later analysis will show that, with a predicate-logical language, the complexity of semantic abduction $\models$ will become higher than that. The reason is that it mixes derivability with non-derivability (because of the consistency condition).

So, our best chance for achieving significant completeness is with an intermediate language, like that of propositional logic. In that case, abduction is still decidable, and we may hope to find simple proof rules for it as well. (Cf. [Tam94] for the technically similar enterprise of completely axiomatizing simultaneous 'proofs' and 'fallacies' in propositional logic.) Can we convert our representation arguments into full-fledged completeness proofs when we add propositional operators $\neg, \wedge, \vee$? We have already seen that we do get natural valid principles like disjunction of antecedents and conjunction of consequents. However, there is no general method that connects a representational result into more familiar propositional completeness arguments. A case of successful (though non-trivial) transfer is in [Kan93], but essential difficulties are identified in [Gro95].

Instead of solving the issue of completeness here, we merely propose the following axioms and rules for a sequent calculus for consistent abduction (which we label as $\models_c$) in what follows:

- Axiom: $p \models_c p$
- Rules for Conjunction:

$$\wedge_1 \frac{\Theta \models_c \varphi_1, \qquad \Theta \models_c \varphi_2}{\Theta \models_c \varphi_1 \wedge \varphi_2}$$

The following are valid provided that α, ψ are formulas with only positive propositional letters:

$$\wedge_2 \frac{\alpha \models_c \alpha \qquad \psi \models_c \psi}{\alpha, \psi \models_c \alpha}$$

$$\wedge_3 \frac{\alpha, \psi \models_c \varphi}{\alpha \wedge \psi \models_c \varphi}$$

- Rules For Disjunction:

$$\vee_1 \frac{\Theta_1 \models_c \varphi \qquad \Theta_2 \models_c \varphi}{\Theta_1 \vee \Theta_2 \models_c \varphi}$$

$$\vee_2 \frac{\Theta \models_c \varphi}{\Theta \models_c \varphi \vee \psi}$$

$$\vee_3 \frac{\Theta \models_c \varphi}{\Theta \models_c \psi \vee \varphi}$$

- Rules for Negation:

$$\neg_1 \frac{\Theta, A \models_c \varphi}{\Theta \models_c \varphi \vee \neg A}$$

$$\neg_2 \frac{\Theta \models_c \varphi \vee A \qquad \Theta \wedge \neg A \models_c \psi}{\Theta \wedge \neg A \models_c \varphi}$$

It is easy to see that these rules are sound on the interpretation of $\models$ as consistent abduction. This calculus is already unlike most usual logical systems, though. First of all there is no substitution rule, as $p \models p$ is an axiom, whereas in general $\psi \not\models \psi$ unless ψ has only positive propositional letters, in which case it is proved to be consistent. By itself, this is not dramatic (for instance, several modal logics exist without a valid substitution rule), but it is certainly uncommon. Moreover, note that the rules which "move things to the left" ($\neg_2$) are different from their classical counterparts, and others ($\wedge_3$) are familiar but here a condition to ensure consistency is added. Even so, one can certainly do practical work with a calculus like this.

For instance, all valid principles of classical propositional logic that do not involve negations are derivable here. Semantically, this makes sense, as positive formulas are always consistent without special precautions. On the other hand, it is easy to check that the calculus provides no proof for a typically invalid sequent like $p \wedge \neg p \models p \wedge \neg p$[8].

[8]The reason is that their cut-free classical proofs (satisfying the subformula property) involve only conjunction and disjunction - for which we have the standard rules.

Digression:
A general semantic view of abductive explanatory consequence

Speaking generally, we can view a ternary inference relation $\Theta \mid \alpha \Rightarrow \varphi$ as a ternary relation C (T, A, F) between sets of models for, respectively, Θ, α, and φ. What structural rules do is constrain these relations to just a subclass of all possibilities. (This type of analysis has analogies with the theory of generalized quantifiers in natural language semantics. It may be found in [vBe84a] on the model theory of verisimilitude, or in [vBe96b] on general consequence relations in the philosophy of science.) When enough rules are imposed we may represent a consequence relation by means of simpler notions, involving only part of the a priori relevant $2^3 = 8$ "regions" of models induced by our three argument sets.

In this light, the earlier representation arguments might even be enhanced by including logical operators. We merely provide an indication. It can be seen easily that, in the presence of disjunction, our explanatory abduction satisfies full Boolean 'Distributivity' for its abducible argument α_i:

$\Theta \mid \bigvee_i \alpha_i \Rightarrow \varphi$ iff for some i, $\Theta \mid \alpha_i \Rightarrow \varphi$.

Principles like this can be used to reduce the complexity of a consequence relation. For instance, the predicate argument A may now be reduced to a point wise one, as any set A is the union of all singletons $\{a\}$ with $a \in$ A.

Complexity

Our next question addresses the complexity of different versions of abduction. Non-monotonic logics may be better than classical ones for modelling common sense reasoning and scientific inquiry. But their gain in expressive power usually comes at the price of higher complexity, and abduction is no exception. Our interest is then to briefly compare the complexity of abduction to that of classical logic. We have no definite results here, but we do have some conjectures. In particular, we look at consistent abduction, beginning with predicate logic.

Predicate-logical validity is undecidable by Church's Theorem. Its exact complexity is Σ^0_1 (the validities are recursively enumerable, but not recursive). (To understand this outcome, think of the equivalent assertion of derivability: "there exists a P: P is a proof for φ".) More generally, Σ (or Π) notation refers to the usual prenex forms for definability of notions in the Arithmetical Hierarchy. Complexity is measured here by looking at the quantifier prenex, followed by a decidable matrix predicate. A subscript n indicates n quantifier changes in the prenex. (If a notion is both Σ_n and Π_n, it is called Δ_n.) The complementary notion of satisfiability is also undecidable, being definable in the form Π^0_1. Now, abductive consequence moves further up in this hierarchy.

In order to show that consistent abduction is not Δ^0_2-complete we have the following. The statement that "Θ, α is consistent" is Π^0_1, while the statement that

"$\Theta, \alpha \models \varphi$" is Σ_1^0 (cf. the above observations). Therefore, their conjunction may be written, using well-known prenex operations, in either of the following forms:

$$\exists\forall DEC \qquad \text{or} \qquad \forall\exists DEC.$$

Hence consistent abduction is in Δ_2^0. This analysis gives an upper bound only. But we cannot do better than this. So it is also a lower bound. For the sake of reductio, suppose that consistent abduction were Σ_1^0. Then we could reduce satisfiability of any formula B effectively to the abductive consequence $B, B \Rightarrow B$, and hence we would have that satisfiability is also Σ_1^0. But then, Post's Theorem says that a notion which is both Σ_1^0 and Π_1^0 must be decidable. This is a contradiction, and hence $\Theta, \alpha \Rightarrow \varphi$ is not Σ_1^0. Likewise, consistent abduction cannot be Π_1^0, because of another reduction: this time from the validity of any formula B to True, True $\Rightarrow B$.

Consistent abduction is not Δ_2^0-complete. Although it is in Δ_2^0 and is not Π_1^0, the latter is not sufficient to prove its hardness and thereby completeness, for that we would have to show that every Δ_2^0 predicate may be reduced to consistent abduction, and we can only prove that it can be written as a conjunction of Σ_1^0 and Π_1^0, showing that it belongs to a relatively simple part of Δ_2^0.

By similar arguments we can show that the earlier weak explanatory abduction is in Δ_2^0 – and the same holds for other variants that we considered. Therefore, our strategy in this chapter of adding amendments to classical consequence is costly, as it increases its complexity. On the other hand, we seem to pay the price just once. It makes no difference with respect to complexity whether we add one or all of the abductive requirements at once. We do not have similar results about the cases with minimality and preference, as their complexity will depend on the complexity of our (unspecified) preference order.

Complexity may be lower in a number of practically important cases. First, consider poorer languages. In particular, for *propositional* logic, all our notions of abduction remain obviously decidable. Nevertheless, their fine-structure will be different. Propositional satisfiability is NP-complete, while validity is Co-NP-complete.

Another direction would restrict attention to useful fragments of predicate logic. For example, universal clauses without function symbols have a decidable consequence problem. Therefore we have the following:

PROPOSITION 4 *All our notions of abductive explanatory inference are decidable over universal clauses.*

Finally, complexity as measured in the above sense may miss out on some good features of abductive reasoning, such as possible natural bounds on search space for abducibles. A very detailed study on the complexity of logic-based abduction which takes into account different kinds of theories (propositional, clausal, Horn) as well as several minimality measures are found in [EG95].

The Role of Language

Our notions of abduction all work for arbitrary formulas, and hence they have no bias toward any special formal language. But in practice, we can often do with simpler forms. E.g., observations φ will often be atoms, and the same holds for explanations α. Here are a few observations showing what may happen.

Syntactic restrictions may make for 'special effects'. For instance, our discussion of minimal abduction contained 'Carnap's trick', which shows that the choice of $\alpha = \Theta \rightarrow \varphi$ will always do for a minimal solution. But notice that this trivialization no longer works when only atomic explanations are allowed.

Here is another example. Let Θ consist of propositional Horn clauses only. In that case, we can determine the minimal abduction for an atomic conclusion directly. A simple example will demonstrate the general method:

Let $\Theta = \{q \wedge r \rightarrow s, p \wedge s \rightarrow q, p \wedge t \rightarrow q\}$ and $\varphi = \{q\}$

$$q \wedge r \rightarrow s, p \wedge s \rightarrow q, p \wedge t \rightarrow q, \alpha? \Rightarrow q$$

(i) $\Theta, \alpha \models ((p \wedge s \rightarrow q) \wedge (p \wedge t \rightarrow q)) \rightarrow q$
(ii) $\Theta, \alpha \models (p \wedge s) \vee (p \wedge t) \vee q$

That is, first make the conjunction of all formulas in Θ having q for head and construct the implication to q (i), obtaining a formula which is already an abductive solution (a slightly simpler form than $\Theta \rightarrow \varphi$).Then construct an equivalent simpler formula (ii) of which each disjunct is also an abductive solution. (Note that one of them is the trivial one). Thus, it is relatively easier to perform this process over a simple theory rather than having to engage in a complicated reasoning process to produce abductive explanations.

Finally, we mention another partly linguistic, partly ontological issue that comes up naturally in abduction. As philosophers of science have observed, there seems to be a natural distinction between 'individual facts' and 'general laws' in explanation. Roughly speaking, the latter belong to the theory Θ, while the former occur as explananda and explanantia. But intuitively, the logical basis for this distinction does not seem to lie in syntax, but rather in the nature of things. How could we make such a distinction? ([Fla95] mentions this issue as one of the major open questions in understanding abduction, and even its implementations.) Here is what we think has to be the way to go. Explanations are sought in some specific situation, where we can check specific facts. Moreover, we adduce general laws, not tied to this situation, which involve general reasoning about the kind of situation that we are in. The latter picture is not what is given to us by classical logic. We would rather have to think of a mixed situation (as in, say, the computer program Tarski's World, cf. [BE93]), where we have two sources of information. One is direct querying of the current situation, the other general deduction (provided that it is sound with respect to this situation.) The proper format for abduction then becomes a mixture of 'theorem proving' and 'model checking' (cf. [SUM96]). Unfortunately, this would go far beyond the bounds of this book.

5. Discussion and Conclusions

Is Abductive Explanatory Inference a Logical System?

The notion of abduction as a logical inference goes back to Peirce's distinction into kinds of logical reasoning, in which abduction plays the role of hypothetical inference. Therefore, its certainty is low and its non-monotonicity high. Though these aspects make it difficult to be handled, it is certainly a logical system of its own kind, which may be classified of the inductive type within Haack's approach[9]. It shares the language with classical logic and abductive conclusions are not valid by only means of the classical consequence (but abduction may have the underlying consequence relation of the deductive type). It can even be somewhat identified with the second characterization of induction, namely in which it is improbable 'supposing'/given that the 'hypotheses'/premisses are true that the conclusion is false, and therefore we can assert that the conclusion is true in a tentative way. But abduction may also be classified as a deviant system, such as in the explanatory abductive version, in which the premisses really contribute for asserting the conclusion, they are relevant in a way to the conclusion. In other words, the language is the same, but the consequence relation is more demanding to the conclusion.

The various types of abductive explanatory styles in a larger universe of other deductive and inductive systems of logic naturally commits us to a pluralistic and a global view of logic, such as Haack's own position, in which there is a variety of logical systems which rather than competing and being rival to each other, they are complementary in that each of them has a specific notion of validity corresponding to an extra-systematic one and a rigorous way for validating arguments, for it makes sense to speak of a logical system as correct or incorrect, having several of them. And finally, the global view states for abduction that it must aspire to global application, irrespective of subject-matter, and thus found in scientific reasoning and in common sense reasoning alike.

Abductive Explanatory Inference as a Structured Logical Inference

Studying abduction as a kind of logical inference has provided much more detail to the broad schema in the previous chapters. Different conditions for a formula to count as a genuine explanation, gave rise to different abductive styles of inference. Moreover, the latter can be used over different underlying

[9] Although Haack does not include explicitly abduction in her classification, she admits her existence [Haa78, page 12n].

notions of consequence (classical, preferential, statistical). The resulting abductive explanatory logics have links with existing proposals in the philosophy of science, and even further back in time, with Bolzano's notion of deducibility. They tend to be non-monotonic in nature. Further logical analysis of some key examples revealed many further structural rules. In particular, consistent abduction was completely characterized. Finally, we have discussed possible complete systems for special kinds of abduction, as well as the complexity of abduction in general.

The analysis of abductive explanatory inference at such an abstract meta-logical level, has allowed for an outlook from a purely structural perspective. We have taken its bare bones and study its consequence type with respect to itself (reflexivity), to the ability to handle new information (monotonicity), to the loss of repeated information (contraction), to the order in which premisses appear (permutation) and to the ability of handling chains of arguments (cut). In short, we tested abductive inference with respect to its ability to react to a changing world. As it turned out, none of the above properties were really an issue for plain abduction, for it is ruled by classical consequence, and therefore observes the same behaviour as that of classical reasoning. It easily allows reflexivity and new information does not invalidate in any way previous one. Moreover, in plain abduction premisse order does not affect the outcome of reasoning. Finally, cut is an easy rule to follow. In contrast, consistent abduction is not even classically reflexive, Still, every formula that in turn explains consistently (or it is explained consistently by something else) is reflexive, and thus ensuring that consistency is preserved. Consistent abduction is also very sensitive to the growth of information, as inconsistent information cannot come in at all. But if the new data explains something else together with the theory, then it is possible to add it as new information. Finally, consistent abduction also handles a somewhat sophisticated kind of cut, giving thus a way to chain the arguments. The only rules, which these two types of abductive reasoning share, are contraction of repeated formulae and permutation.

Here is what we consider the main outcomes of our analysis. We can see abductive explanatory inference as a more structured form of consequence, whose behavior is different from classical logic, but which still has clear inferential structure. The modifications of classical structural rules, which arise in this process, may even be of interest by themselves – and we see this whole area as a new challenge to logicians. Note that we did not locate the 'logical' character of abduction in any specific set of (modified) structural rules. If pressed, we would say that some modified versions of Reflexivity, Monotonicity and Cut seem essential – but we have not been able to find a single formulation that would stand once and for all. (Cf. [Gab94a] and [Gab94b] for a fuller discussion of the latter point.) Another noteworthy point was our ternary format of inference, which gives different roles to the theory and explanation on the one hand, and to the conclusion on the other. This leads to finer-grained views of inference rules, whose interest has been demonstrated.

Abductive Explanatory Inference and Geometries

Nevertheless, the structural characterization we have proposed still leaves a question unanswered, namely, in what sense an structural characterization leads to a logic, to a full syntactic or semantic characterization. Even more, despite the technical results presented, some readers may still doubt whether abductive reasoning can be considered really *logical*, perhaps it is more appropriate to render it as a special type of reasoning. After all, by accepting abductive reasoning as logical we are accepting a system that only produces tentative conclusions and not certainties as it is the case for classical reasoning. Let me precise these questions as follows:

1 In what sense the structural characterization of consistent abduction does lead to its logic?

2 Are non-classical inferences, such as abduction, really *logical*?

Regarding the first of these questions, its answer concerns a mathematical technical problem. That is, it implies a reformulation of the representation theorem into a completeness theorem, for a logical language without operators (recall that structural rules are pure, they have no connectives). Furthermore, a syntactic characterization of abduction requires the extension of the logical language, to include axioms and operators in order to formulate rules with connectives and so construct an adequate logical abductive calculi. This way to proceed, which is to obtain a syntax out of an structural characterization, has been explored with success for other logics, such as dynamic, relevance and categorial. Regarding a semantics for abduction, there is also some exploratory work in this direction, using an extended version of semantic tableaux (cf. next chapter). However, I conjecture that an abductive version such as the one allowing all conditions at once does produce a logic that is incomplete.

Regarding the second question, its answer concerns a terminological question of what we want to denote by the term *logic*. Although structural analysis of consequence has proved very fruitful and has even been proposed as a distinguished enterprise of *Descriptive Logic* in [Fla95], many logicians remain doubtful, and withhold the status of bona fide ‘logical inference’ to the products of non-standard styles.

This situation is somewhat reminiscent of the emergence of non-euclidean geometries in the nineteenth century. Euclidean geometry was thought of as the one and only geometry until the fifth postulate (the parallel axiom) was rejected, giving rise to new geometries. Most prominently, the one by Lobachevsky , which admits of more than one parallel, and the one by Riemann admitting none. The legitimacy of these geometries was initially doubted but their impact gradually emerged[10]. In our context, it is not geometry but styles of reasoning that occupy the space, and there is not one postulate under critical scrutiny, but several. Rejecting monotonicity gives rise to the family of non-monotonic logics, and rejecting permutation leads to styles of dynamic inference. Linear logics on the other hand, are created by rejecting contraction. All these alternative logics might get their empirical vindication, too – as reflecting different *modes* of human reasoning.

Whether non-classical modes of reasoning are really logical is like asking if non-euclidean geometries are really geometries. The issue is largely terminological, and we might decide – as Quine did on another occasion (cf.[Qui61]) – to *just* give conservatives the word 'logic' for the more narrowly described variety, using the word 'reasoning' or some other suitable substitute for the wider brands. In any case, an analysis in terms of structural rules does help us to bring to light interesting features of abduction, logical or not.

Conclusions

Summarizing, we have shown that abduction can be studied with profit as a purely logical notion of inference. Of course, we have not exhausted this viewpoint here – but we must leave its full exploration to other logicians. Also, we do not claim that this analysis exhausts all essential features of abduction, as discussed in chapter 2. To the contrary, there are clear limitations to what our present perspective can achieve. While we were successful in characterizing what an explanation is, and even show how it should behave inferentially under addition or deletion of information, the generation of abductive explanations was not discussed at all. The latter procedural enterprise is the topic of our next chapter. Another clear limitation is our restriction to the case of 'novelty', where there is no conflict between the theory and the observation. For the case of 'anomaly', we need to go into theory revision, as will happen in chapter 8. That chapter will also resume some threads from the present one, including a full version of abduction, in which all our cumulative conditions are incorporated. The latter will be needed for our discussion of Hempel's deductive-nomological model of explanation.

Related Work

Abduction has been recognized as a non-monotonic logic but with few exceptions, no study has been made to characterize it as a logical inference. In [Kon90] a general theory of abduction is defined as classical inference with

[10]The analogy with logic can be carried even further, as these new geometries were sometimes labeled 'meta-geometries'.

the additional conditions of consistency and minimality, and it is proved to be implied by Reiter's causal theories [Rei87], in which a diagnosis is a minimal set of abnormalities that is consistent with the observed behaviour of a system. Abduction is also proposed as a procedural mechanism in which the input Q is an abductive stimulus (goal), and we are interested in Δ' such that $\Delta + \Delta'$ explains Q (with some suitable underlying inference) [Gab94b, p. 199].

Another approach, closer to our own, though developed independently, is found in Peter Flach's PhD dissertation "Conjectures: an inquiry concerning the logic of induction" [Fla95], which we will now briefly describe and compare to our work (some of what follows is based on a more recent version of his proposal [Fla96a].)

Flach's logic of induction

Flach's thesis is concerned with a logical study of conjectural reasoning, complemented with an application to relational databases. An inductive consequence relation $\prec$ ($\prec \subseteq LxL$, L a propositional language) is a set of formulae; $\alpha \prec \beta$ interpreted as "β is a possible inductive hypothesis that *explains* α", or as: "β is a possible inductive hypothesis *confirmed by* α". The main reason for this distinction is to dissolve the paradoxical situation posed by Hempel's adequacy conditions for confirmatory reasoning [Hem43, Hem45], namely that in which a piece of evidence E could confirm any hypothesis whatsoever[11]. Therefore, two systems are proposed: one for the logic of confirmation and the other for the logic of explanation, each one provided with an appropriate representation theorem for its characterization. These two systems share a set of inductive principles and differ mainly in that explanations may be strengthened without ceasing to be explanations (H5), and confirmed hypotheses may be weakened without being disconfirmed (H2). To give an idea of the kind of principles these systems share, we show two of them, the well-known principles of verification and falsification in Philosophy of Science:

I1 If $\alpha \prec \beta$ and $\models \alpha \wedge \beta \rightarrow \gamma$, then $\alpha \wedge \gamma \prec \beta$.

I2 If $\alpha \prec \beta$ and $\models \alpha \wedge \beta \rightarrow \gamma$, then $\alpha \wedge \neg\gamma \not\prec \beta$.

They state that when a hypothesis β is tentatively concluded on the basis of evidence α, and a prediction γ drawn from α and β is observed, then β counts as a hypothesis for both α and γ (I1), and not for α and $\neg\gamma$ (I2) (a consequence of the latter is that reflexivity is only valid for consistent formulae).

[11]This situation arises from accepting reflexivity (H1: any observation report is confirmed by itself) and stating on the one hand that if an observation report confirms a hypothesis, then it also confirms every consequence of it (H2), and on the other that if an observation report confirms a hypothesis, then it also confirms every formula logically entailing it (H5).

Comparison to our work

Despite differences in notation and terminology, Flach's approach is connected to ours in several ways. Its philosophical motivation is based on Peirce and Hempel, its methodology is also based on structural rules, and we agree that the relationship between explananda and explanandum is a logical parameter (rather than fixed to deduction) and on the need for complementing the logical approach with a computational perspective. Once we get into the details however, our proposals present some fundamental differences, from a philosophical as well as a logical point of view.

Flach departs from Hempel's work on confirmation [Hem43, Hem45], while ours is based on later proposals on explanation [HO48, Hem65]. This leads to a discrepancy in our basic principles. One example is (consistent) reflexivity; a general inductive principle for Flach but rejected by us for explanatory abduction (since one of Hempel's explanatory adequacy conditions imply that it is invalid, cf. chapter 5). Note that this property reflects a more fundamental difference between confirmation and explanation than H2 and H5: evidence confirms itself, but it does not explain itself [12]. There are also differences in the technical setup of our systems. Although Flach's notion of inductive reasoning may be viewed as a strengthened form of logical entailment, the representation of the additional conditions is explicit in the rules rather than within the consequence relation. Nevertheless, there are interesting analogies between the two approaches, which we must leave to future work. We conclude with a general remark. A salient point in both our approaches is the importance of consistency, also crucial in Hempel's adequacy conditions both for confirmation and explanation, and in AI approaches to abduction. Thus, Bolzano's notion of deducibility comes back as capturing an intrinsic property of conjectural reasoning in general.

[12] Flach correctly points out that Hempel's own solution to the paradox was to drop condition (H5) from his logic of confirmation. Our observation is that the fact that Hempel later developed an independent account for the logic of explanation [HO48, Hem65], suggests he clearly separated confirmation from explanation. In fact his logic for the latter differs in more principles than the ones mentioned above.