Seeking Explanations:

Abduction in Logic, Philosophy of Science and Artifical Intelligence

Atocha Aliseda-LLera

Seeking Explanations:

Abduction in Logic, Philosophy of Science and Artifical Intelligence

ILLC Dissertation Series 1997-4

For further information about ILLC-publications, please contact

Institute for Logic, Language and Computation
Universiteit van Amsterdam
Plantage Muidergracht 24
1018 TV Amsterdam
phone: +31-20-5256090
fax: +31-20-5255101
e-mail: illc@wins.uva.nl

SEEKING EXPLANATIONS: ABDUCTION IN LOGIC, PHILOSOPHY OF SCIENCE AND ARTIFICIAL INTELLIGENCE

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF PHILOSOPHY
INTERDEPARTMENTAL PROGRAM IN PHILOSOPHY AND SYMBOLIC SYSTEMS
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

By
Atocha Aliseda-LLera
August 1997

Promotor: Prof.dr. J. van Benthem
Faculteit Wiskunde en Informatica
Universiteit van Amsterdam
Plantage Muidergracht 24
1018 TV Amsterdam

The investigations were supported by the Universidad Nacional Autònoma de México, Instituto de Investigaciones Filosóficas.

ISBN: 90–74795–73–0

Contents

Acknowledgements

It is a privilege to have five professors on my reading committee representing the different areas of my Ph.D. program in Philosophy and Symbolic Systems: Computer Science, Linguistics, Logic, Philosophy, and Psychology.

Tom Wasow was the first person to point me in the direction of abduction. He gave me useful comments on earlier versions of this dissertation, always insisting that it be readable to non experts. It was also a pleasure to work with him this last year coordinating the undergraduate Symbolic Systems program.

Dagfinn Føllesdal encouraged me to continue exploring connections between abduction and philosophy of science. Yoav Shoham and Pat Suppes gave me very good advice about future expansions of this work. Jim Greeno chaired my defense and also gave me helpful suggestions. For their help with my dissertation, and for the classes in which they taught me, I am very grateful.

To my advisor, Johan van Benthem, I offer my deepest gratitude. He is a wonderful teacher and mentor. Among many things, Johan taught me that basic notions in logic may give rise to the most interesting questions. Working with him meant amazement and challenge, and his passion for logic is an endless inspiration to my work.

I received help from numerous other people and institutions. For inspiring conversations, invaluable bibliography, and constructive criticism in early stages of this project, I wish to thank: Carmina Curcó, Pablo Gervás, Dinda Gorlée, Jerry Hobbs, Marianne Kalsbeek, Geert-Jan Kruijff, Ralf Müller, David Pearce, and Víctor Sánchez-Valencia. I would also like to thank the organizers and participants of the 'ECAI'96 Workshop on Abductive and Inductive Reasoning' to whom I had the opportunity to present a small part of this dissertation. Moreover, I thank the students in the senior seminar on 'Explanatory Reasoning' which I led last winter, for challenging many ideas of my favorite authors.

I gladly acknowledge the financial help I received from the Universidad Nacional Autónoma de México (UNAM), which sponsored my education. The former director of the Institute of Philosophical Research (IIF), León Olivé, supported and approved my project, and the current director, Olbeth Hansberg, helped me above and beyond the initial commitment.

I had a wonderful office at the Center for the Study of Language and Information (CSLI). I thank the director John Perry for that privilege, and especially for allowing me to use the speech recognition system of his Archimedes project when I suffered an injury caused by extensive typing. CSLI is housed in two buildings with unusual Spanish names. I am happy to say that after all these years of working in Cordura ("prudence"), I had my defense in Ventura ("fortune").

On a more personal note, my situation called for extensive traveling to London, where my husband was doing his Ph.D. at the London School of Economics. This involved many people who kept their home open for us to stay, and their basements empty for our multiple boxes. They provided us with their love, support, and many pleasant *sobremesas* (after dinner conversations). Thanks a million to our parents and siblings, to Aida and Martín in Mexico, to Vivian and Eduardo in Oakland, to Marty in Palo Alto, to Julienne in Amsterdam, and specially, to Rosie in London. I would also like to mention my friends (you know who you are) who kept me sane and happy while at planet Stanford and abroad.

Finally, I would like to mention the key people who motivated me to pursue a graduate degree in the United States. I thank Adolfo García de la Sienra, for encouraging me to come to Stanford. His intellectual generosity sent me to my first international congress to present our work. I also thank Carlos Torres Alcaráz, who introduced me to the beautiful framework of semantic tableaux and directed my undergraduate thesis, and Elisa Viso Gurovich, my first computer science teacher, who has given me her unconditional support all along. I would also like to express my gratitude to my late father, who inculcated in me the will to learn and the satisfaction of doing what you like. He would be very happy to know that I finally finished school!

Above all, I thank Rodolfo, who shares with me his passion for music and his adventurous spirit. He kept all too well his promise to take care of me through graduate school, even though he had to write his own thesis, and live in another country. I test many of my ideas with him, and I hope we will some day write that book we have talked about. I am delighted that we complement our twenty years of friendship with nine of marriage. This dissertation is, of course, dedicated to him with all my love.

Chapter 1

What is Abduction? Overview and Proposal for Investigation

1.1 What is Abduction?

A central theme in the study of human reasoning is the construction of explanations that give us an understanding of the world we live in. Broadly speaking, *abduction* is a reasoning process invoked to explain a puzzling observation. A typical example is a practical competence like medical diagnosis. When a doctor observes a symptom in a patient, she hypothesizes about its possible causes, based on her knowledge of the causal relations between diseases and symptoms. This is a practical setting. Abduction also occurs in more theoretical scientific contexts. For instance, it has been claimed [Han61],[CP, 2.623] that when Kepler discovered that Mars had an elliptical orbit, his reasoning was abductive. But, abduction may also be found in our day-to-day common sense reasoning. If we wake up, and the lawn is wet, we might explain this observation by assuming that it must have rained, or by assuming that the sprinklers have been on. Abduction is thinking from evidence to explanation, a type of reasoning characteristic of many different situations with incomplete information.

The history of this type of reasoning goes back to Antiquity. It has been compared with Aristotle's *apagoge* [CP, 2.776,5.144] which intended to capture a non-strictly deductive type of reasoning whose conclusions are not necessary, but merely possible (not to be confused with *epagoge*, the Aristotelian term for induction). Later on, abduction as reasoning from effects to causes is extensively discussed in Laplace's famous memoirs [Lap04, Sup96] as an important methodology in the sciences. In the modern age, this reasoning was put on the intellectual agenda under the name 'abduction' by C.S. Peirce [CP, 5.189].

To study a type of reasoning that occurs in contexts as varied as scientific discovery, medical diagnosis, and common sense, suitably broad features must be provided, that cover a lot of cases, and yet leave some significant substance to the notion of abduction. The purpose of this preliminary chapter is to introduce these, which will lead to the more specific questions treated in subsequent chapters. But before we start with a more general analysis, let us expand our stock of examples.

Examples

The term 'abduction' is used in the literature for a variety of explanatory processes. We list a few, partly to show what we must cover, and partly, to show what we will leave aside.

1. Common Sense: Explaining observations with simple facts.

> All you know is that the lawn gets wet either when it rains, or when the sprinklers are on. You wake up in the morning and notice that the lawn is wet. Therefore you hypothesize that it rained during the night or that the sprinklers had been on.

2. Common Sense: Laying causal connections between facts.

> You observe that a certain type of clouds (nimbostratus) usually precede rainfall. You see those clouds from your window at night. Next morning you see that the lawn is wet. Therefore, you infer a causal connection between the nimbostratus at night, and the lawn being wet.

3. Common Sense: Facing a Contradiction.

You know that rain causes the lawn to get wet, and that it is indeed raining. However, you observe that the lawn is not wet. How could you explain this anomaly?

4. Statistical Reasoning: Medical Diagnosis[1].

Jane Jones recovered rapidly from a streptococcus infection after she was given a dose of penicillin. Almost all streptococcus infections clear up quickly upon administration of penicillin, unless they are penicillin-resistant, in which case the probability of quick recovery is rather small. The doctor knew that Jane's infection is of the penicillin-resistant type, and is completely puzzled by her recovery. Jane Jones then confesses that her grandmother had given her Belladonna, a homeopatic medicine which stimulates the immune system by strengthening the physiological resources of the patient to fight infectious diseases.

The examples so far are fairly typical of what our later analysis can deal with. But actual explanatory reasoning can be more complicated than this. For instance, even in common sense settings, there may be various options, which are considered in some sequence, depending on your memory and 'computation strategy'.

5. Common Sense: When something does not work.

You come into your house late at night, and notice that the light in your room, which is always left on, is off. It has being raining very heavily, and so you think some power line went down, but the lights in the rest of the house work fine. Then, you wonder if you left both heaters on, something which usually causes the breakers to cut off, so you check them: but they are OK. Finally, a simpler explanation crosses your mind. Maybe the light bulb of your lamp which you last saw working well, is worn out, and needs replacing.

[1]This is an adaptation of Hempel's famous illustration of his Inductive-Statistical model of explanation as presented in [Sal92]. The part about homeopathy is entirely mine, however.

So, abduction involves computation over various candidates, depending on your background knowledge. In a scientific setting, this means that abductions will depend on the relevant background theory, as well as one's methodological 'working habits'. We mention one often-cited example, even though we should state clearly at this point that it goes far beyond what we shall eventually deal with in our analysis.

6. Scientific Reasoning: Kepler's discovery[2].

> One of Johannes Kepler's great discoveries was that the orbit of the planets is elliptical rather than circular. What initially led to this discovery was his observation that the longitudes of Mars did not fit circular orbits. However, before even dreaming that the best explanation involved ellipses instead of circles, he tried several other forms. Moreover, Kepler had to make several other assumptions about the planetary system, without which his discovery does not work. His heliocentric view allowed him to think that the sun, so near to the center of the planetary system, and so large, must somehow cause the planets to move as they do. In addition to this strong conjecture, he also had to generalize his findings for Mars to all planets, by assuming that the same physical conditions obtained throughout the solar system. This whole process of explanation took many years.

It will be clear that the Kepler example has a loose end, so to speak. How we construct the explanation depends on what we take to be his scientific background theory. This is a general feature of abductions: explanation is always explanation w.r.t. some body of beliefs. But even this is not the only parameter that plays a role. One could multiply the above examples, and find still further complicating factors. Sometimes, no single obvious explanation is available, but rather several competing ones - and we have to select. Sometimes, the explanation involves not just advancing facts or rules in our current conceptual frame, but rather the creation of new concepts, that allow for new description of the relevant phenomena. Evidently, we must draw a line somewhere in our present study.

[2] This example is a simplification of one in [Han61].

All our examples were instances of reasoning in which an explanation is needed to account for a certain phenomenon. Is there more unity than this? At first glance, the only clear common feature is that these are not cases of ordinary deductive reasoning, and this for a number of reasons. In particular, the explanations produced might be *defeated*. Maybe the lawn is wet because children have been playing with water. Co-occurrence of clouds and the lawn being wet does not necessarily link them in a causal way. Jane's recovery might after all be due to a normal process of the body. What we learn subsequently can invalidate an earlier abductive conclusion. Moreover, the reasoning involved in these examples seems to go in reverse to ordinary deduction, as all these cases run from evidence to hypothesis, and not from data to conclusion, as it is usual in deductive patterns. Finally, describing the way in which an explanation is found, does not seem to follow specific rules. Indeed, the precise nature of Kepler's 'discovery' remains under intensive debate[3].

What we would like to do is the following. Standard deductive logic cannot account for the above types of reasoning. In this century, under the influence of foundational research in mathematics, there has been a contraction of concerns in logic to this deductive core. The result was a loss in breadth, but also a clear gain in depth. By now, an impressive body of results has been obtained about deduction - and we would like to study the wider field of abduction while hanging on to these standards of rigor and clarity. Of course, we cannot do everything at once, and achieve the whole agenda of logic in its traditional open-ended form. We want to find some features of abduction that allow for concrete logical analysis, thereby extending the scope of standard methods. In the next section, we discuss three main features, that occur across all of the above examples (properly viewed), which will be important in our investigation.

[3]For Peirce, Kepler's reasoning was a prime piece of abduction [CP, 1.71,2.96], whereas for Mill it was merely a description of the facts [Mill 58, Bk III, ch II. 3], [CP, 1.71–4]. Even nowadays one finds very different reconstructions. While Hanson presents Kepler's heliocentric view as an essential assumption [Han61], Thagard thinks he could make the discovery assuming instead that the earth was stationary and the sun moves around it [Tha92]. Still a different account of how this discovery can be made is given in [SLB81, LSB87].

Three Faces of Abduction

We shall now introduce three broad oppositions that help in clarifying what abduction is about. At the end, we indicate how these will be dealt with in this thesis.

Abduction: Product or Process?

The logical key words of *judgment* and *proof* are nouns which denote either an activity, indicated by their corresponding verb, or the result of that activity. In just the same way, the common word *explanation* – which we treat as largely synonymous with abduction – may be used both to refer to a finished product, the explanation of a phenomenon, or to an activity, the process that led to that explanation. These two uses are closely related. The process of explanation produces explanations as its product, but the two are not the same.

One can relate this distinction to more traditional ones. An example is Reichenbach's [Rei38] well-known opposition of 'context of discovery' versus 'context of justification'. Kepler's explanation-product "the orbit of the planets is elliptical", which justifies the observed facts, does not include the explanation-process of how he came to make this discovery. The context of discovery has often been taken to be purely psychological, but this does not preclude its exact study. Cognitive psychologists study mental patterns of discovery, learning theorists in AI study formal hypothesis formation, and one can even work with concrete computational algorithms that produce explanations. To be sure, it is a matter of debate whether Kepler's reasoning may be modeled by a computer. (For a computer program that claims to model this particular discovery, cf. [SLB81].) However this may be, one can certainly write simple programs that produce common sense explanations of 'why the lawn is wet', as we will show later on. On the other hand, once produced, explanations are public objects of 'justification', that can be checked and tested by independent logical criteria.

The product–process distinction has been recognized by logicians [Bet59, vBe93], in the context of deductive reasoning, as well as by philosophers of science [Rub90, Sal90] in the context of scientific explanation. Both lead to interesting questions by themselves, and so does their interplay. Likewise, these two faces of abduction are both relevant for our study. On the product side, our focus will be on conditions that give a piece of information explanatory force, and on the process side, we will be concerned with the design of algorithms that produce explanations.

Abduction: Construction or Selection?

Given a fact to be explained, there are often several possible explanations, but only one (or a few) that counts as the best one. Pending subsequent testing, in our common sense example of light failure, several explanations account for the unexpected darkness of the room (power line down, breakers cut off, bulb worn out). But only one may be considered as 'best' explaining the event, namely the one that really happened. But other preference criteria may be appropriate, too, especially when we have no direct test available.

Thus, abduction is connected to both hypothesis construction and hypothesis selection. Some authors consider these processes as two separate steps, construction dealing with what counts as a possible explanation, and selection with applying some preference criterion over possible explanations to select the best one. Other authors regard abduction as a single process by which a single best explanation is constructed. Our position is an intermediate one. We will split abduction into a first phase of hypothesis construction, but also acknowledge a next phase of hypothesis selection. We shall mainly focus on a characterization of possible explanations. We will argue that the notion of a 'best explanation' necessarily involves contextual aspects, varying from application to application. There is at least a new parameter of preference ranking here. Although there exist both a philosophical tradition on the logic of preference [Wri63], and logical systems in AI for handling preferences that may be used to single out best explanations[Sho88, DP91b], the resulting study would take us too far afield.

Abduction vs Induction

Once beyond deductive logic, diverse terminologies are being used. Perhaps the most widely used term is inductive reasoning [Mill 58, Sal90, HHN86, Tha88, Fla95, Mic94].

Abduction is another focus, and it is important, at least, to clarify its relationship to induction. For C.S. Peirce, as we shall see, 'deduction', 'induction' and 'abduction' formed a natural triangle – but the literature in general shows many overlaps, and even confusions.

Since the time of John Stuart Mill (1806-1873), the technical name given to all kinds of non-deductive reasoning has been 'induction', though several *methods for discovery and demonstration of causal relationships* [Mill 58] were recognized. These included generalizing from a sample to a general property, and reasoning from data to a causal hypothesis (the latter further divided into methods of agreement, difference, residues, and concomitant variation). A more refined and modern terminology is 'enumerative induction' and 'explanatory induction', of which 'inductive generalization', 'inductive projection', 'statistical syllogism', 'concept formation' are some instances. Such a broad connotation of the term induction continues to the present day. For instance, in the computational philosophy of science, induction is understood *"in the broad sense of any kind of inference that expands knowledge in the face of uncertainty* [Tha88].

Another 'heavy term' for non-deductive reasoning is *statistical reasoning*, introducing a probabilistic flavour, like our example of medical diagnosis, in which possible explanations are not certain but only probable. Statistical reasoning exhibits the same diversity as abduction. First of all, just as the latter is strongly identified with *backwards deduction* (as we shall see later on in this chapter), the former finds its 'reverse notion' in probability (The problem in probability is: given an stochastic model, what can we say about the outcomes? The problem is statistics is the reverse: given a set of outcomes, what can we say about the model?) Both abduction and statistical reasoning are closely linked with notions like confirmation (the testing of hypothesis) and likelihood (a measure for alternative hypotheses).

On the other hand, some authors take induction as an instance of abduction. Abduction as *inference to the best explanation* is considered by Harman [Har65] as the basic form of non-deductive inference, which includes (enumerative) induction as a special case.

This confusion returns in artificial intelligence. 'Induction' is used for the process of learning from examples – but also for creating a theory to explain the observed facts [Sha91], thus making abduction an instance of induction. Abduction is usually restricted to producing explanations in the form of facts. When the explanations are rules, it is regarded as part of induction. The relationship between abduction and induction (properly defined) has been the topic for recent workshops in AI conferences [ECAI96].

To clear up all these conflicts, one might want to coin new terminology altogether. Many authors write as if there were pre-ordained, reasonably clear notions of abduction and its rivals, which we only have to analyze to get a clear picture. But these technical terms may be irretrievably confused in their full generality, burdened with the debris of defunct philosophical theories. Therefore, I have argued for a new term of *"explanatory reasoning"* in [Ali96a], trying to describe its fundamental aspects without having to decide if they are instances of either abduction or induction. In this broader perspective, we can also capture explanation for more than one instance or for generalizations, – which we have not mentioned at all – and introduce further fine–structure. For example, given two observed events, in order to find an explanation that accounts for them, it must be decided whether they are causally connected (eg. entering the copier room and the lights going on), correlated with a common cause (eg. observing both the barometric pressure and the temperature dropping at the same time), or just coincidental without any link (you reading this paragraph in place A while I revise it somewhere in place B). But in this dissertation, we shall concentrate on explanatory reasoning from simple facts, giving us enough variety for now. Hopefully, this case study of abduction will lead to broader clarity of definition as well.

More precisely, we shall understand abduction as reasoning from *a single observation to its explanations*, and induction as *enumerative induction* from samples to general statements. While induction explains a set of observations, abduction explains a single one. Induction makes a prediction for further observations, abduction does not (directly) account for later observations. While induction needs no background theory per se, abduction relies on a background theory to construct and test its explanations. (Note that this abductive formulation does not commit ourselves to any specific logical inference, kind of observation, or form of explanation.)

As for their similarities, induction and abduction are both non-monotonic types of inference, and both run in opposite direction to standard deduction. In non-monotonic inference, new premises may invalidate a previous valid argument. In the terminology of philosophers of science, non-monotonic inferences are not *erosion proof* [Sal92]. Qua direction, induction and abduction both run from evidence to explanation. In logical terms, this may be viewed as going from a conclusion to (part of) its premises, in reverse of ordinary deduction. We will return to these issues in much more detail in our logical chapter 2.

1.2 The Founding Father: C.S. Peirce

The literature on abduction is so vast, that we cannot undertake a complete survey here. What we shall do is survey some highlights, starting with the historical sources of the modern use of the term. In this field, all 20th century roads lead back to the work of C.S. Peirce. For a much more elaborate historical analysis cf.[Ali95]. Together with the other sources to be discussed, the coming sections will lead up to further parameters for the general taxonomy of abduction that we propose toward the end of this chapter.

Understanding Peirce's Abduction

Charles Sanders Peirce (1839-1914), the founder of American pragmatism was the first philosopher to give to abduction a logical form, and hence his relevance to our study. However, his notion of abduction is a difficult one to unravel. On the one hand, it is entangled with many other aspects of his philosophy, and on the other hand, several different conceptions of abduction evolved in his thought. We will point out a few general aspects of his theory of inquiry, and later concentrate on some of its more logical aspects.

The notions of logical inference and of validity that Peirce puts forward go beyond our present understanding of what logic is about. They are linked to his epistemology, a dynamic view of thought as logical inquiry, and correspond to a deep philosophical concern, that of studying the nature of synthetic reasoning.

In his early theory Peirce proposed three modes of reasoning: deduction, induction, and abduction, each of which corresponds to a syllogistic form, illustrated by the following, often quoted example [CP, 2.623]:

> DEDUCTION
>
> Rule.– All the beans from this bag are white.
> Case.– These beans are from this bag.
> Result.– These beans are white.
>
> INDUCTION
>
> Case.– These beans are from this bag.
> Result.– These beans are white.
> Rule.– All the beans from this bag are white.
>
> ABDUCTION
>
> Rule.– All the beans from this bag are white.
> Result.– These beans are white.
> Case.– These beans are from this bag.

Of these, deduction is the only reasoning which is completely certain, infering its 'Result' as a necessary conclusion. Induction produces a 'Rule' validated only in the 'long run' [CP, 5.170], and abduction merely suggests that something may be 'the Case' [CP, 5.171].

Later on, Peirce proposed these types of reasoning as the stages composing a method for logical inquiry, of which abduction is the beginning:

> *"From its [abductive] suggestion deduction can draw a prediction which can be tested by induction."* [CP, 5.171].

Abduction plays a role in direct perceptual judgments, in which:

> *"The abductive suggestion comes to us as a flash"* [CP, 5.181]

As well as in the general process of invention:

> *"It [abduction] is the only logical operation which introduces any new ideas"* [CP, 5.171].

In all this, abduction is both *"an act of insight and an inference"* as has been suggested in [And86]. These explications do not fix one unique notion. Peirce refined his views on abduction throughout his work. He first identified abduction with the syllogistic form above, to later enrich this idea by the more general conception of:

> "*the process of forming an explanatory hypothesis*" [CP, 5.171].

And also refering to it as:

> *"The process of choosing a hypothesis"* [CP, 7.219]

Something which suggests that he did not always distinguish clearly between the construction and the selection of a hypothesis – as was pointed out in [Fan70]. The evolution of his theory is also reflected in the varied terminology he used to refer to abduction; beginning with *presumption* and *hypothesis* [CP, 2.776,2.623], then using *abduction* and *retroduction* interchangeably [CP, 1.68,2.776,7.97].

A nice concise account of the development of abduction in Peirce, which clearly distinguishes three stages in the evolution of his thought is given in [Fan70]. Another key reference on Peirce's abduction, in its relation to creativity in art and science is found in [And87].

The Key Features of Peircean Abduction

For Peirce, three aspects determine whether a hypothesis is promising: it must be *explanatory*, *testable*, and *economic*. A hypothesis is an explanation if it accounts for the facts. Its status is that of a suggestion until it is verified, which explains the need for the testability criterion. Finally, the motivation for the economic criterion is twofold: a response to the practical problem of having innumerable explanatory hypotheses to test, as well as the need for a criterion to select the best explanation amongst the testable ones.

For the explanatory aspect, Pierce gave the following often-quoted logical formulation [CP, 5.189]:

> The surprising fact, C, is observed.
> But if A were true, C would be a matter of course.
> Hence, there is reason to suspect that A is true.

This formulation has played a key role in Peirce scholarship, and it has been the point of departure of recent studies on abductive reasoning in artificial intelligence [KKT95, HSAM90, PG87]. However, no one seems to agree on its interpretation. We will also give our own (for details, cf. [Ali95]).

Interpreting Peirce's Abduction

The notion of abduction in Peirce has puzzled scholars all along. Some have concluded that Peirce held no coherent view on abduction at all [Fra58], others have tried to give a joint account with induction [Rei70] and still others claim it is a form of inverted modus ponens [And86]. A more modern view is found in [Kap90] who interprets Peirce's abduction as a form of heuristics. An account that tries to make sense of the two extremes of abduction, both as a guessing instinct and as a rational activity is found in [Ayi74]. I have argued in [Ali95] that this diversity suggests that Peirce recognized not only different types of reasoning, but also several degrees within each one, and even merges between the types. In the context of perception he writes:

> "The perceptual judgements, are to be regarded as extreme cases of abductive inferences" [CP, 5.181]

Abductory induction, on the other hand, is suggested when some kind of guess work is involved in the reasoning [CP, 6.526]. Anderson [And87] also recognizes several degrees in Peirce's notion of creativity.

This multiplicity returns in artificial intelligence. [Fla96b] suggests that some confusions in modern accounts of abduction in AI can be traced back to Peirce's two theories of abduction: the earlier syllogistic one and the later inferential one. As to more general semiotic aspects of Peirce's philosophy, another proposal for characterizing abduction in AI is found in [Kru95].

Our own understanding of abductive reasoning reflects this Peircean diversity in part, taking abduction as a style of logical reasoning that occurs at different levels and in several degrees. These will be reflected in our proposal for a taxonomy with several 'parameters' for abductive reasoning. This scheme will be richer than the logical form for Peirce's abductive formulation often encountered in the literature:

$$\begin{array}{l} C \\ \underline{A \rightarrow C} \\ A \end{array}$$

where the status of A is that of a tentative explanation. Though simple and intuitive, this formulation captures neither the fact that C is surprising nor the additional aspects of testability and economy that Peirce proposed. For instance, existing computational accounts of abduction do not capture that C is a surprising fact. In our logical treatment of abduction in subsequent chapters, we bring out at least two aspects of Peirce's formulation that go beyond the preceding schema, namely: 1) C is a surprising fact, 2) A is an explanation. For further aspects, we refer to subsequent sections. The additional criteria of testability and economy are not part of our framework. Testability as understood by Peirce is an extra-logical empirical criterion, while economy concerns the selection of explanations, which we already put aside as a further stage of abduction requiring a separate study.

1.3 Philosophy of Science

Peirce's work stands at the crossroads of many traditions, including logic, epistemology, and philosophy of science. Especially, the latter field has continued many of his central concerns. Abduction is clearly akin to core notions of modern methodology, such as explanation, induction, discovery, and heuristics. We have already discussed a connection between process-product aspects of abduction and the well-known division between contexts of discovery and justification [Rei38]. We shall discuss several further points of contact in chapter 4 below. But for the moment, we start with a first foray.

The 'Received View': explanation as a product

The dominant trend in philosophy has focused on abduction as product rather than a process, just as it has done for other epistemic notions. Aristotle, Mill, and in this century, the influential philosopher of science Carl Hempel, all based their accounts of explanation on proposing criteria to characterize its products. These accounts generally classify into argumentative and non-argumentative types of explanation [Rub90, Sal90, Nag79]. Of particular importance to us is the 'argumentative' Hempelian tradition. Its followers aim to model empirical why-questions, whose answers are scientific explanations in the form of arguments. In these arguments, the 'explanandum' (the fact to be explained) is derived (deductively or inductively) from the 'explananda' (that which does the explaining) supplemented with relevant 'laws' (general or statistical) and 'initial conditions'. For instance, the fact that an explosion occurred may be explained by my lighting the match, given the laws of physics, and initial conditions to the effect that oxygen was present, the match was not wet, et cetera.

In its deductive version, the Hempelian account, found in many standard texts on the philosophy of science [Nag79, Sal90] is called *deductive-nomological*, for obvious reasons. But its engine is not just standard deduction. Additional restrictions must be imposed on the relation between explananda and explanantia, as neither deduction nor induction is a sufficient condition for genuine explanation. To mention a simple example, every formula is derivable from itself ($\varphi \vdash \varphi$), but it seems counterintuitive, or at least very uninformative, to explain anything by itself.

Other, non-deductive approaches to explanation exist in the literature. For instance, [Rub90] points at these two:

> [Sal77, p.159] takes them to be: *"an assemblage of factors that are statistically relevant ..."*
>
> while [vFr80, p.134] makes them simply: *"an answer"*.

For Salmon, the question is not how probable the explanans renders the explanandum, but rather whether the facts adduced make a difference to the probability of the explanandum. Moreover, this relationship need not be in the form of an argument. For van Fraassen, a representative of pragmatic approaches to explanation, the explanandum is a contrastive why-question. Thus, rather than asking "why φ?", one asks "why φ rather than γ?". The pragmatic view seems closer to abduction as a process, and indeed, the focus on questions introduces some dynamics of explaining. Still, it does not tell us how to produce explanations.

There are also alternative deductive approaches. An example is the work of Rescher [Res78], which introduces a *direction of thought*. Interestingly, this establishes a temporal distinction between 'prediction' and 'retroduction' (Rescher's term for abduction), by marking the precedence of the explanandum over the hypothesis in the latter case. Another, and quite famous deductivist tradition is Popper's logic of scientific discovery [Pop58]. Its method of conjectures and refutations proposes the testing of hypotheses, by attempting to refute them:

> *"The actual procedure of science is to operate with conjectures: to jump to conclusions – often after one single observation"* [Pop63, p.53].
>
> "*Thus science starts from problems, and not from observations; though observations may give rise to a problem, specially if they are unexpected; that is to say, if they clash with our expectations or theories*". [Pop63, p.222].

Popper's deductive focus is on refutation of falsehoods, rather than explanation of truths. One might speculate about a similar 'negative' counterpart to abduction. Although Popper's method claims to be a logic of scientific discovery, he views the actual construction of explanations as an exclusive matter for psychology – and hence his 'trial and error' philosophy offers no further clues for our next topic.

What is common to all these approaches in the philosophy of science is the importance of a hidden parameter in abduction. Whether with Hempel, Salmon, or Popper, scientific explanation never takes place in isolation, but always in the context of some *background theory*. This additional parameter will become part of our general scheme to be proposed below.

The 'Neglected View': explanation as a process

Much more marginal in the philosophy of science are accounts of explanation that focus on explanatory processes as such. One early author emphasizing explanation as a process of discovery is Hanson ([Han61]), who gave an account of patterns of discovery, recognizing a central role for abduction (which he calls 'retroduction'). Also relevant here is the work by Lakatos ([Lak76]), a critical response to Popper's logic of scientific discovery. For Lakatos, there is only a fallibilistic logic of discovery, which is neither psychology nor logic, but an independent discipline, the *logic of heuristic*. He pays particular attention to processes that created new concepts in mathematics – often referring to Polya ([Pol45]) as the founding father of heuristics in mathematical discovery[4]. We will come back to this issue later in the chapter, when presenting further fields of application.

What these examples reveal is that in science, explanation involves the invention of new concepts, just as much as the positing of new statements (in some fixed conceptual framework). So far, this has not led to extensive formal studies of concept formation, similar to what is known about deductive logic. (Exceptions that prove the rule are occasional uses of Beth's Definability Theorem in the philosophical literature. A similar lacuna vis-a-vis concept revision exists in the current theory of belief revision in AI.)

Thus, philosophy of science offers little for our interests in abductive processes. We are certainly in sympathy with the demand for conceptual change in explanation – but this topic will remain beyond the technical scope of this thesis.

[4]In fact Polya contrasts two types of arguments. A demonstrative syllogism in which from $A \rightarrow B$, and B false, $\neg A$ is concluded, and a *heuristic syllogism* in which from $A \rightarrow B$, and B true, it follows that A is more credible. The latter, of course, recalls Peirce's abductive formulation.

1.4 Artificial Intelligence

Our next area of comparison is artificial intelligence. The transition with the previous section is less abrupt than it may be seem. It has often been noted, by looking at the respective research agendas, that artificial intelligence is philosophy of science, pursued by other means (cf. [Tan92]). Research on abductive reasoning in AI dates back to 1973 [Pop73], but it is only fairly recently that it has attracted great interest, in areas like logic programming [KKT95], knowledge assimilation [KM94], and diagnosis [PG87], to name just a few. Abduction is also coming up in the context of data bases and knowledge bases, that is, in mainstream computer science.

In this setting, the product-process distinction has a natural counterpart, namely, in logic-based vs computational-based approaches to abduction. While the former focuses on giving a semantics to the logic of abduction, usually defined as 'backwards deduction plus additional conditions', the latter is concerned with providing algorithms to produce abductions.

It is impossible to give an overview here of this exploding field. Therefore, we limit ourselves to (1) a brief description of abduction as logical inference, (2) a presentation of abduction in logic programming, and (3) a sketch of the relevance of abductive thinking in knowledge representation. There is much more in this field of potential philosophical interest, however. For abduction in bayesian networks, connectionism, and many other angles, the reader is advised to consult [Jos94, PR90, Kon96, Pau93, AAA90].

Abduction as Logical Inference

The general trend in logic based approaches to abduction in AI interprets abduction as *backwards deduction plus additional conditions*. This brings it very close to deductive-nomological explanation in the Hempel style, witness the following format. What follows is the standard version of abduction as deduction via some consistent additional assumption, satisfying certain extra conditions. It combines some common requirements from the literature (cf. [Kon90, KKT95, MP93] and chapter 2 for further motivation):

Given a theory Θ (a set of formulae) and a formula φ (an atomic formula), α is an explanation if

1. $\Theta \cup \alpha \models \varphi$
2. α is consistent with Θ
3. α is 'minimal' (there are several ways to characterize minimality, to be discussed in chapter 2).
4. α has some restricted syntactical form (usually an atomic formula or a conjunction of them).

An additional condition not always made explicit is that $\Theta \not\models \varphi$. This says that the fact to be explained should not already follow from the background theory alone. Sometimes, the latter condition figures as a precondition for an *abductive problem.*

What can one say in general about the properties of such an 'enriched' notion of consequence? As we have mentioned before, a new logical trend in AI studies variations of classical consequence via their 'structural rules', which govern the combination of basic inferences, without referring to any special logical connectives. (Cf. the analysis of non-monotonic consequence relations in AI of [Gab94a], [KLM90], and the analysis of dynamic styles of inference in linguistics and cognition in [vBe90].) Perhaps the first example of this approach in abduction is the work in [Fla95] – and indeed our analysis in chapter 2 will follow this same pattern.

Abduction in Logic Programming

Logic Programming [LLo87, Kow79] was introduced by Kowalski and Colmerauer in 1974, and is implemented as (amongst others) the programming language Prolog. It is inspired by first-order logic, and it consists of logic programs, queries, and a underlying inferential mechanism known as resolution[5].

[5] Roughly speaking, a Prolog program P is an ordered set of rules and facts. Rules are restricted to horn-clause form $A \leftarrow L_1, \ldots, L_n$ in which each L_i is either an atom A_i or its negation $\neg A_i$. A query q (theorem) is posed to program P to be solved (proved). If the query follows from the program, a positive answer is produced, and so the query is said to be succesful. Otherwise, a

Abduction emerges naturally in logic programming as a 'repair mechanism', completing a program with the facts needed for a query to succeed. This may be illustrated by our rain example (1) from the introduction in Prolog:

Program P**:**

lawn-wet $\leftarrow$ rain.

lawn-wet $\leftarrow$ sprinklers-on.

Query q: lawn-wet.

Given program P, query q does not succeed because it is not derivable from the program. For q to succeed, either one (or all) of the facts 'rain', 'sprinklers-on', 'lawn-wet' would have to be added to the program. Abduction is the process by which these additional facts are produced. This is done via an extension of the resolution mechanism that comes into play when the backtracking mechanism fails. In our example above, instead of declaring failure when either of the above facts is not found in the program, they are marked as 'hypothesis', and proposed as those formulas which, if added to the program, would make the query succeed.

In actual Prolog abduction, for these facts to be counted as abductions, they have to belong to a pre-defined set of 'abducibles', and to be verified by additional conditions (so-called 'integrity constraints'), in order to prevent a combinatorial explosion of possible explanations.

In logic programming, the procedure for constructing explanations is left entirely to the resolution mechanism, which affects not only the order in which the possible explanations are produced, but also restricts the form of explanations. Notice that rules cannot occur as abducibles, since explanations are produced out of sub-goal literals that fail during the backtracking mechanism. Therefore, our common sense

---

negative answer is produced, indicating that the query has failed. However, the interpretation of negation is 'by failure'. That is, 'no' means 'it is not derivable from the available information in P' – without implying that the negation of the query $\neg q$ is derivable instead. Resolution is an inferential mechanism based on refutation working backwards: from the negation of the query to the data in the program. In the course of this process, valuable by-products appear: the so-called 'computed answer substitutions', which give more detailed information on the objects satisfying given queries.

example (2) in which a causal connection is abduced to explain why the lawn is wet, cannot be implemented in logic programming[6]. The additional restrictions select the best hypothesis. Thus, processes of both construction and selection of explanations are clearly marked in logic programming . (Another relevant connection here is to recent research in 'inductive logic programming' [Mic94], which integrates abduction and induction.)

Logic programming does not use blind deduction. Different control mechanisms for proof search determine how queries are processed. This additional degree of freedom is crucial to the efficiency of the enterprise. Hence, different control policies will vary in the abductions produced, their form and the order in which they appear. To us, this variety suggests that the procedural notion of abduction is intensional, and must be identified with different practices, rather than with one deterministic fixed procedure.

Abduction and Theories of Epistemic Change

Most of the logic-based and computation-based approaches to abduction reviewed in the preceding sections assume that neither the explanandum nor its negation is derivable from the background theory ($\Theta \not\models \varphi$, $\Theta \not\models \neg\varphi$). This leaves no room to represent problems like our common sense light example (5) in which the theory expects the contrary of our observation. (Namely, that the light in my room is on.) These are cases where the theory needs to be *revised* in order to account for the observation. Such cases arise in practical settings like diagnostic reasoning [PR90], belief revision in databases [AD94] and theory refinement in machine learning [SL90, Gin88].

When importing revision into abductive reasoning, an obvious related territory is theories of belief change in AI. Mostly inspired by the work of Gärdenfors [Gär88] (a work whose roots lie in the philosophy of science), these theories describe how to incorporate a new piece of information into a database, a scientific theory, or a set

[6]At least, this is how the implementation of abduction in logic programming stands as of now. It is of course possible to write extended programs that produce these type of explanations.

of common sense beliefs. The three main types of belief change are operations of 'expansion', 'contraction', and 'revision'. A theory may be expanded by adding new formulas, contracted by deleting existing formulas, or revised by first being contracted and then expanded. These operations are defined in such a way as to ensure that the theory or belief system remains consistent and suitably 'closed' when incorporating new information.

Our earlier cases of abduction may be described now as expansions, where the background theory gets extended to account for a new fact. What is added are cases where the surprising observation (in Peirce's sense) calls for revision. Either way, this perspective highlights the essential role of the background theory in explanation. Belief revision theories provide an explicit calculus of modification for both cases. It must be clarified however, that changes occur only in the theory, as the situation or world to be modelled is supposed to be static, only new information is coming in. Another important type of epistemic change studied in AI is that of *update*, the process of keeping beliefs up–to–date as the world changes. We will not analyze this second process here – even though we are confident that it can be done in the same style proposed here. Evidently, all this ties in very neatly with our earlier findings. (For instance, the theories involved in abductive belief revision might be structured like those provided by our discussion, or by cues from the philosophy of science.) We will explore this connection in more detail in chapter 4.

1.5 Further Fields of Application

The above survey is by no means exhaustive. Abduction occurs in many other research areas, of which we will mention three: linguistics, cognitive science, and mathematics (the former was indeed an early motivation of this dissertation). Although we will not pursue these lines elsewhere in this dissertation, they do provide a better perspective on the broader interest of our topic. For instance, abduction in cognitive science is an interdisciplinary theme relating about all areas relevant to a Ph.D. in Symbolic Systems.

Abduction in Linguistics

In linguistics, abduction has been proposed as a process for natural language interpretation [HSAM90], where our 'observations' are the utterances that we hear (or read). More precisely, interpreting a sentence in discourse is viewed as providing a best explanation of why the sentence would be true. For instance, a listener or reader abduces assumptions in order to resolve references for definite descriptions ("the cat is on the mat" invites you to assume that there is a cat and a mat), and dynamically accommodates them as presupposed information for the sentence being heard or read.

Abduction also finds a place in theories of language acquisition. Most prominently, Chomsky proposed that learning a language is a process of theory construction. A child 'abduces' the rules of grammar guided by her innate knowledge of language universals. Indeed in [Cho72], he refers to Peirce's justification for the logic of abduction, – based on the human capacity for 'guesing the right hypotheses', to reinforce his claim that language acquisition from highly restricted data is possible.

Abduction has also been used in the semantics of *questions*. Questions are then the input to abductive procedures generating answers to them. Some work has been done in this direction in natural language as a mechanism for dealing with indirect replies to yes-no questions [GJK94]. Of course, the most obvious case where abduction is explicitly called for are "Why" questions, inviting the other person to provide a reason or cause.

Abduction also connects up with linguistic *presuppositions*, which are progressively accommodated during a discourse. [Ali93] treats accommodation as a non-monotonic process, in which presuppositions are not direct updates for explicit utterances, but rather abductions that can be refuted because of later information. Accommodation can be described as a *repair strategy* in which the presuppositions to be accommodated are not part of the background. In fact, the linguistic literature has finer views of types of accommodation (cf. the 'local'/'global' distinction in [Hei83]), which might correspond to the two abductive 'triggers' proposed in the next section. A broader study on presuppositions which considers abductive mechanisms and uses the framework of semantic tableaux to represent the context of discourse, is found in [Ger95].

More generally, we feel that the taxonomy proposed later in this chapter might correlate with the linguistic diversity of presupposition (triggered by definite descriptions, verbs, adverbs, et cetera) – but we must leave this as an open question.

Abduction in Cognitive Science

In cognitive science, abduction is a crucial ingredient in processes like inference, learning, and discovery, performed by people to build theories of the world surrounding them. There is a growing literature on computer programs modeling these processes (cf. [HHN86, Tha92, SL90, Gie91]).

An important pilot reference is Simon [SLB81], whose authors claim that scientific discovery can be viewed as a problem-solving activity. Although there is no precise method by which scientific discovery is achieved, as a form of problem solving, it can be pursued via several methodologies. The authors distinguish between weak and strong methods of discovery. The former is the type of problem solving used in novel domains. It is characterized by its generality, since it does not require in-depth knowledge of its particular domain. In contrast, strong methods are used for cases in which our domain knowledge is rich, and are specially designed for one specific structure.

Weak methods include generation, testing, heuristic methods, and means-ends analysis, to build explanations and solutions for given problems. These methods have proved useful in AI and cognitive simulation, and are used by several programs. An example is the BACON system which models explanations and descriptive scientific laws, such as Kepler's law, Ohm's law, etcetera. It is a matter of debate if BACON really makes discoveries, since it produces theories new to the program but not to the world, and its discoveries seem spoonfed rather than created. Be that as it may, our analysis in this dissertation does not claim to model these higher types of explanations and discoveries.

Another noteworthy reference is found in the new field of 'computational philosophy of science' ([Tha88]), and in broader computational cognitive studies of inductive reasoning ([HHN86]). These studies distinguish several relevant processes: simple abduction, existential abduction, abduction to rules, and analogical abduction. We will propose a similar multiplicity in what follows.

Abduction in Mathematics

Abduction in mathematics is usually identified with notions like discovery and heuristics. A key reference in this area is the work by the earlier mentioned G. Polya [Pol45, Pol54, Pol62]. In the context of number theory, for example, a general property may be guessed by observing some relation as in:

$$3+7=10, \qquad 3+17=20, \qquad 13+17=30$$

Notice that the numbers 3,7,13,17 are all odd primes and that the sum of any of two of them is an even number. An initial observation of this kind eventually led Goldbach (with the help of Euler) to formulate his famous conjecture: *'Every even number greater than two is the sum of two odd primes'*.

Another example is found in a configuration of numbers, such as in the well–known Pascal's triangle [Pol62]:

$$\begin{array}{ccccccccc} & & & 1 & & & & = & 1 \\ & & 1 & & 1 & & & = & 2 \\ & 1 & & 2 & & 1 & & = & 4 \\ 1 & & 3 & & 3 & & 1 & = & 8 \end{array}$$

There are several 'hidden' properties in this triangle, which the reader may or may not discover depending on her previous training and mathematical knowledge. A simple one is that any number different from 1 is the sum of two other numbers in the array, namely of its northwest and northeast neighbors (eg. $3 = 1 + 2$). A more complex relationship is this: the numbers in any row sum to a power of 2. More precisely,

$$\binom{n}{0} + \cdots + \binom{n}{n} = 2^n$$

See [Pol62] for more details on 'abducing' laws about the binomial coefficients in the Pascal Triangle.

The next step is to ask why these properties hold, and then proceed to prove them. Goldbach's conjecture remains unsolved (it has not been possible to prove it or to refute it); it has only been verified for a large number of cases (the latest news is that it is true for all integers less than 4.10^{11}, cf. [Rib96]). The results regarding Pascal's triangle on the other hand, have many different proofs, depending one's particular taste and knoweldge of geometrical, recursive, and computational methods. (Cf. [Pol62] for a detailed discussion of alternative proofs.)

According to Polya, a mathematician discovers just as a naturalist does, by observing the collection of his specimens (numbers or birds) and then guessing their connection and relationship [Pol54, p.47]. However, the two differ in that verification by observation for the naturalist is enough, whereas the mathematician requires a proof to accept her findings. This points to a unique feature of mathematics: once a theorem finds a proof, it cannot be defeated. Thus, mathematical truths are eternal, with possibly many ways of being *explained*. On the other hand, some findings may remain unexplained forever. Abduction in mathematics shows very well that observing goes beyond visual perception, as familiarity with the field is required to find 'surprising facts'. Moreover, the relationship between observation and proof need not be causal, it is just pure mathematical structure that links them together.

Much more complex cases of mathematical discovery can be studied, in which concept formation is involved. A recent and interesting approach along these lines is found in [Vis97], which proposes a catalogue of procedures for creating concepts when solving problems. These include 'redescription', 'substitution', and 'transposition', which are explicitly related to Peirce's treatment of abduction.

1.6 A Taxonomy for Abduction

What we have seen so far may be summarized as follows. Abduction is a general process of explanation, whose products are specific explanations, with a certain inferential structure. We consider these two aspects of equal importance. Moreover, on the process side, we distinguished between constructing possible explanations and selecting the best one amongst these. This thesis is mainly concerned with the structure of explanations as products, and with the process of constructing these.

As for the logical form of abduction, we have found that it may be viewed as a threefold relation:

$$\Theta, \alpha \Rightarrow \varphi$$

between an observation φ, an abduced item α, and a background theory Θ. (Other parameters are possible here, such as a preference ranking - but these would rather concern the further selection process.) Against this background, we propose three main parameters that determine types of abduction. (i) An 'inferential parameter' ($\Rightarrow$) sets some suitable logical relationship among explananda, background theory, and explanandum. (ii) Next, 'triggers' determine what kind of abduction is to be performed: φ may be a novel phenomenon, or it may be in conflict with the theory Θ. (iii) Finally, 'outcomes' (α) are the various products of an abductive process: facts, rules, or even new theories.

Abductive Parameters

Varieties of Inference

In the above schema, the notion of explanatory inference $\Rightarrow$ is not fixed. It can be classical derivability $\vdash$ or semantic entailment $\models$, but it does not have to be. Instead, we regard it as a parameter which can be set independently. It ranges over such diverse values as probable inference ($\Theta, \alpha \Rightarrow_{probable} \varphi$), in which the explanandum renders the explanandum only highly probable, or as the inferential mechanism of logic programming ($\Theta, \alpha \Rightarrow_{prolog} \varphi$). Further interpretations include dynamic inference ($\Theta, \alpha \Rightarrow_{dynamic} \varphi$, cf. [vBe96a]), replacing truth by information change potential along the lines of belief update or revision. Our point here is that abduction is not one specific non-standard logical inference mechanism, but rather a way of using any one of these.

Different Triggers

According to Peirce, as we saw, abductive reasoning is triggered by a *surprising phenomenon*. The notion of surprise, however, is a relative one, for a fact φ is surprising only with respect to some backgound theory Θ providing 'expectations'. What is surprising to me (eg. that the lights go on as I enter the copier room) might not be surprising to you. We interpret a surprising fact as one which needs an explanation. From a logical point of view, this assumes that the fact is not already explained by the background theory Θ: $\Theta \not\Rightarrow \varphi$.

Moreover, our claim is that one also needs to consider the status of the negation of φ. Does the theory explain the negation of observation instead ($\Theta \Rightarrow \neg\varphi$)? Thus, we identify at least two triggers for abduction: *novelty* and *anomaly*:

- Abductive Novelty: $\Theta \not\Rightarrow \varphi$, $\Theta \not\Rightarrow \neg\varphi$

  φ is novel. It cannot be explained ($\Theta \not\Rightarrow \varphi$), but it is consistent with the theory ($\Theta \not\Rightarrow \neg\varphi$).

- Abductive Anomaly: $\Theta \not\Rightarrow \varphi$, $\Theta \Rightarrow \neg\varphi$.

  φ is anomalous. The theory explains rather its negation ($\Theta \Rightarrow \neg\varphi$).

In the computational literature on abduction, novelty is the condition for an *abductive problem* [KKT95]. My suggestion is to incorporate anomaly as a second basic type.

Of course, non-surprising facts (where $\Theta \Rightarrow \varphi$) should not be candidates for explanation. Even so, one might speculate if facts which are merely probable on the basis of Θ might still need explanation of some sort to further cement their status.

Different Outcomes

Abducibles themselves come in various forms: facts, rules, or even theories. Sometimes one simple fact suffices to explain a surprising phenomenon, such as rain explaining why the lawn is wet. In other cases, a rule establishing a causal connection might serve as an explanation, as in our case connecting cloud types with rainfall. And many cases of abduction in science provide new theories to explain surprising facts. These different options may sometimes exist for the same observation, depending on how seriously we want to take it. In this thesis, we shall mainly consider explanations in the forms of atomic facts, conjunctions of them and simple conditionals, but we do make occasional excursions to more complex kinds of statements.

Moreover, we are aware of the fact that genuine explanations sometimes introduce new concepts, over and above the given vocabulary. (For instance, th eventual explanation of planetary motion was not Kepler's, but Newton's, who introduced a new notion of 'force' – and then derived elliptic motion via the Law of Gravity.) Except for passing references in subsequent chapters, abduction via new concepts will be outside the scope of our analysis.

Abductive Processes

Once the above parameters get set, several kinds of abductive processes arise. For example, abduction triggered by novelty with an underlying deductive inference, calls for a process by which the theory is expanded with an explanation. The fact to be explained is consistent with the theory, so an explanation added to the theory accounts deductively for the fact. However, when the underlying inference is statistical, in a case of novelty, theory expansion might not be enough. The added statement might lead to a 'marginally consistent' theory with low probability, which would not yield a strong explanation for the observed fact. In such a case, theory revision is needed (ie. removing some data from the theory) to account for the observed fact with high probability. (For a specific example of this latter case cf. chapter 4.)

Our aim is not to classify abductive processes, but rather to point out that several kinds of these are used for different combinations of the above parameters. In the coming chapters we explore in detail some procedures for computing different types of outcomes in a deductive format; those triggered by novelty (chapter 3) and those triggered by anomaly (chapter 4).

Examples Revisited

Given our taxonomy for abductive reasoning, we can now see more patterns across our earlier list of examples. Varying the inferential parameter, we cover not only cases of deduction but also statistical inferences. Thus, Hempel's statistical model of explanation [Hem65] also becomes a case of abduction. Our example (4) of medical diagnosis (Jane's quick recovery) was an instance. Logic programming inference seems more appropriate to an example like (1) (whose overall structure is the similar to (4)). As for triggers, novelty drives both rain examples (1) and (2), as well as the medical diagnosis one (4). A trigger by anomaly occurs in example (3), where the theory predicts the contrary of our observation, the lights-off example (5), and the Kepler example (6), since his initial observation of the longitudes of Mars contradicted the previous rule of circular orbits of the planets. As for different outcomes, examples (1), (3), (4) and (5) abduce facts, examples (2) and (6) produce rules as their forms of explanantia. Different forms of outcomes will play a role in different types of procedures for producing explanations. In computer science jargon, triggers and outcomes are, respectively, preconditions and outputs of abductive devices, whether these be computational procedures or inferential ones.

This taxonomy gives us the big picture of abductive reasoning. In the remainder of this thesis, we are going to investigate various of its aspects, which give rise to more specific logical and computational questions. (Indeed, more than we have been able to answer!) Before embarking upon this course, however, we need to discuss one more general strategic issue, which explains the separation of concerns between chapters 2 and 3 that are to follow.

Abductive Logic: Inference + Search Strategy

Classical logical systems have two components: a semantics and a proof theory. The former aims at characterizing what it is for a formula to be true in a model, and it is based on the notions of truth and interpretation. The latter characterizes what counts as a valid proof for a formula, by providing the inference rules of the system; having for its main notions proof and derivability. These two formats can be studied independently, but they are closely connected. At least in classical (first-order) logic, the completeness theorem tells us that all true formulas have a proof, and vice versa. Many logical systems have been proposed that follow this pattern: propositional logic, predicate logic, modal logic, and various typed logics.

From a modern perspective, however, there is much more to reasoning than this. Computer science has posed a new challenge to logic; that of providing automatic procedures to operate logical systems. This requires a further fine-structure of reasoning. In fact, recent studies in AI give precedence to a *control strategy* for a logic over its complete proof theory. In particular, the heart of logic programming lies in its control strategies, which lead to much greater sensitivity as to the order in which premises are given, the avoidance of search loops, or the possibility to cut proof searches (using the extra-logical operator !) when a solution has been found. These features are extralogical from a classical perspective, but they do have a clear formal structure, which can be brought out, and has independent interest as a formal model for broader patterns of argumentation (cf. [vBe92, Kal95, Kow91]).

Several contemporary authors stress the importance of control strategies, and a more finely-structured algorithmic description of logics. This concern is found both in the logical tradition ([Gab94b, vBe90]), and in the philosophical tradition ([Gil96]), the latter arguing for a conception of logic as: *inference + control*. (Note the shift here away from Kowalski's famous dictum "Algorithm = Logic + Control".) In line with this philosophy, we wish to approach abduction with two things in mind. First, there is the inference parameter, already discussed, which may have several interpretations. But given any specific choice, there is still a significant issue of a suitable search strategy over this inference, which models some particular abductive practice. The former parameter may be defined in semantic or proof-theoretic terms. The search procedure, on the other hand, deals with concrete mechanisms for producing valid inferences. It is then possible to control which kinds of outcome are produced with a certain efficiency. In particular, in abduction, we may want to produce only 'useful' or 'interesting' formulas, preferably even just some 'minimal set' of these.

In this light, the aim of an abductive search procedure is not necessarily completeness with respect to some semantics. A procedure that generates all possible explanations might be of no practical use, and might also miss important features of human abductive activity. In chapter 3, we are going to experiment with semantic tableaux as a vehicle for attractive abductive strategies that can be controlled in various ways.

1.7 Thesis Aim and Overview

The main aim in this dissertation is to study the notion of abduction, that is, reasoning from an observation to its possible explanations, from a logical point of view. This approach to abductive reasoning naturally leads to connections with theories of explanation in the philosophy of science, and to computationally oriented theories of belief change in Artificial Intelligence.

Many different approaches to abduction can be found in the literature, as well as a bewildering variety of instances of explanatory reasoning. To delineate our subject more precisely, and create some order, a general taxonomy for abductive reasoning was proposed in this chapter. Several forms of abduction are obtained by instantiating three parameters: the kind of reasoning involved (e.g., deductive, statistical), the kind of observation triggering the abduction (novelty, or anomaly w.r.t. some background theory), and the kind of explanations produced (facts, rules, or theories). In chapter 2, I choose a number of major variants of abduction, thus conceived, and investigate their logical properties. A convenient measure for this purpose are so-called 'structural rules' of inference. Abduction deviates from classical consequence in this respect, much like many current non-monotonic consequence relations and dynamic styles of inference. As a result we can classify forms of abduction by different structural rules. A more computational analysis of processes producing abductive inferences is then presented in chapter 3, using the framework of semantic tableaux. I show how to implement various search strategies to generate various forms of abductive explanations. Our eventual conclusion is that abductive processes should be our primary concern, with abductive inferences their secondary 'products'. Finally, chapter 4 is a confrontation of the previous analysis with existing themes in the philosophy of science and artificial intelligence. In particular, I analyse two well-known models for scientific explanation (the deductive-nomological one, and the inductive-statistical one) as forms of abduction. This then provides them with a structural logical analysis in the style of chapter 2. Moreover, I argue that abduction can model dynamics of belief revision in artificial intelligence. For this purpose, an extended version of the semantic tableaux of chapter 3 provides a new representation of the operations of expansion and contraction.

Chapter 2

Abduction as Logical Inference

2.1 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'.

The outcomes that we obtain in this first systematic chapter, naturally divided in three parts, are as follows. In the first part (sections 1–5), we propose a general logical format for abduction, involving more parameters than in standard inference, allowing for genuinely different roles of premises. We find a number of natural styles of abduction, rather than one single candidate. These abductive versions are classified by different *structural rules* of inference, and this issue occupies the second part (sections 6–7). As a small contribution to the logical literature in the field, we give a complete characterization of one simple style of abduction, which may also be viewed as the first structural characterization of a natural style of explanation in the philosophy of science. In the third part of this chapter (sections 8–9), 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. Despite these useful insights, pure logical analysis does not exhaust all there is to abduction. In particular, its 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 abduction, which gives us a view of its variety and complexity, and which allows us to raise some interesting new questions.

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 premises P.*

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 premises to conclusions: $P \Rightarrow C$ and the conclusions generated are undefeasible. We briefly recall some features that make abduction a form of inference which 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 premises (or part of them) are the output of the inferential process: $P \Leftarrow C$. Moreover, the *abduced* premise 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 premises 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 artifical intelligence on various notions of common sense reasoning [McC80, Sho88, Gab94a]. 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.

2.2 Directions in Reasoning: Forward and Backward

Intuitively, a valid inference from, say, premises P_1, P_2 to a conclusion C allows for various directions of thought. In a forward direction, given the premises, 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 $P_1 \wedge P_2$, 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 premises 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 premise 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 eventualy failed analyses of possible counterexamples to an inference, read in another they are Gentzen-style sequent derivations of the inference. (We shall be using tableaux in our next chapter, on computing abduction.) Beth himself took this as a formal model of the historical opposition between methods of 'analysis' and 'synthesis' in the development of scientific argument. 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 premises that would support the conclusion[1].

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 1, 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 3. 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.

[1] In this case, a corresponding refutation would rather be a forward process: if the abduced premise 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.

2.3 Formats of Inference: Premises and Background Theory

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

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

These are 'local steps', which take place in the context of some, implicit or explicit, background theory (as we have seen in chapter 1). In this standard format, the background theory is either omitted, or lumped together with the other premises. 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 premise, 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 premises and implicit *background assumptions*. More drastically, premise 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 premises play the role of either scientific laws, or of initial conditions, or of specific explanatory items, suggesting the following format:

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

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 [Gab94a].) While in Hempel's account, structuring the premises makes sure that scientific explanation involves an interplay of laws and facts, Ryan's motivation is resolution of conflicts between premises 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 4. A working version of these ideas is a recent 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 premises 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 premises 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 premises 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 juridicial procedures, which are richer than pure mathematical proof. Toulmin's format of reasoning contains the necessary tags for these procedures:

$$\begin{array}{cccc} \text{Data} & \xrightarrow{\hspace{12em}} & \begin{array}{c}\text{Qualifier}\\ \text{Rebuttal}\end{array} & \text{Claim} \\ & \begin{array}{c}\text{Warrant}\\ | \\ \text{Backing}\end{array} & & \end{array}$$

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) which 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 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 1, 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 inference, in the remainder of this thesis, 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 1 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.

2.4 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 premises 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 which revolve around truth, statistical inference requires not total inclusion of premise 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 update-to-test-consequence:

"process the succesive 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 deafult 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 abduction. On our view, abduction is not a new notion of inference. It is rather a topic-dependent practice of explanatory 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 premises forming *compatible propositions*, those for which some set of ideas make all propositions true when uniformly substitued 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 premises 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 premise 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 4).

2.5 Requirements for Abductive Inference

In this section we define abduction 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 (eg. $r \wedge \neg w$). 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 '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 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 explanation α is only *necessary*, if φ is not already entailed by Θ. Otherwise, any consistent formula will count as an 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 which 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 simplicty in a logical way is as follows:

> Weakest Explanation:
>
> α is the weakest 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 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[2], or even more ambitiously, following current approaches to 'verisimilitude' in the philosophy of science (cf. [Kui87]).

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 Styles

Following our presentation of various requirements for abductive reasoning, we make things more concrete for further reference. We consider five versions of abduction: plain, consistent, explanatory, minimal and preferential, defined as follows:

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

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

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 explanation.

Preferential :

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

(ii) α is the best 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 4, 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 premises 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.

Our more systematic analysis of different abductive styles uses a logical methodology that has recently become popular across a range of non-standard logics.

2.6 Styles of Inference and Structural Rules

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 premises ('monotonicity'), or that we may safely chain valid inferences ('transitivity' or 'cut'). This type of analysis (started in [Sco71]) describes a style of inference at a very abstract structural level, giving its pure combinatorics. It has proved very successful in artifical intelligence for studying different types of plausible reasoning ([KLM90]), and indeed as a general framework for non-monotonic consequence relations ([Gab85]). A new area where it has proved itself is dynamic semantics, where not one but many new notions of dynamic consequences are to be analyzed ([vBe96a]).

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 premises to the left, and one conclusion to the right of the sequent arrow.

Classical Inference

The structural rules for classical inference are the following:

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

- **Contraction:**

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

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

- **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 premise implies itself, no trouble is caused by deleting repeated premises; premises may be permuted without altering validity, adding new information does not invalidate previous conclusions, and premises may be replaced by sequences of premises implying them. In all, these rules allow us to treat the premises 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. Thus, one rule may fit many logics: propositional, first-order, modal, type-theoretic, etc. This makes them different from inference rules like, say, Conjunction of Consequents or Disjunction of Antecedents, which also fix the meaning of conjnction and disjunction.

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) premises in the (models for the) conclusion:

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

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:

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

[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 premises of an argument as a set; Reflexivity is needed to show that for maximal consistent sets, membership and derivability coincide".

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.

Non-Classical Inference

For non-classical consequences, classical structural rules may fail. A well-known example 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 satisfy a weaker form of monotonicity. Additions to the premises are allowed only when these premisses imply them:

- Cautious Monotonicity:

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

Dynamic inference is non-monotonic (inserting arbitrary new processes into a premise sequence can disrupt earlier effects). But it also quarrels with other classical structural rules, such as Cut. But again, representation theorems exist. Thus, the earlier dynamic style known of 'update-to-test' 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 relevant logic, linear logic, and categorial logics (cf. [DH93, vBe91]).

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 on in connection with abduction.

We will look at these matters for abduction in a moment. But, since this perspective may still be unfamiliar to many readers, we provide a small excursion.

Are non-classical inferences really logical?

Structural analysis of consequence relations goes back to Bolzano's program of charting different styles of inference. It has even been proposed as a distinguished enterprise of *Descriptive Logic* in [Fla95]. However, 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[5]. In our context, it is not geometry but styles of reasoning that occupy the space, and there is

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

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.

2.7 Structural Rules For Abduction

In this section we provide structural rules for different versions of abduction with classical consequence underneath. Plain abduction is characterized by classical inference. A complete characterization for consistent abduction is provided. For the explanatory and preferential versions, we just give some structural rules and speculate about their complete characterization.

2.7.1 Consistent Abduction

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 which 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 premises 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 the 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 $\emptyset \subset A_1^* \cap \ldots \cap A_1^* \subseteq B^*$.

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, $A_1, \ldots, A_k \in \bigcap A_i^*$, for each i with $1 \leq i \leq k$, which gives the proper inclusion. Next, let U be any sequence in $\bigcap A_i^*, 1 \leq i \leq k$. 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, $U \in \bigcap A_i^*$, $1 \leq i \leq k$. 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 which 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 premise 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:

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

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

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

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

- Permutation

$$\cfrac{\cfrac{\cfrac{X, A, B, Y \Rightarrow C}{X, A, B, Y \Rightarrow X, A, B, Y \text{ separately}}\,1}{X, B, A, Y \Rightarrow X, B, A, Y \text{ separately}}\,3 \qquad X, A, B, Y \Rightarrow C}{X, B, A, Y \Rightarrow C}\,2$$

- Contraction (one sample case)

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

Thus, consistent abduction defined as classical consequence plus the consistency requirement has appropriate forms of reflexivity, monotonicity, and cut for which it is assured that the premises 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 premise 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 4).

2.7.2 Explanatory Abduction

Explanatory abduction was defined as plain abduction ($\Theta, \alpha \Rightarrow \varphi$) plus two conditions of necessity ($\Theta \not\Rightarrow \varphi$) and insufficiency ($\alpha \not\Rightarrow \varphi$). 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 premises set to achieve a conclusion. It will be be analyzed further in chapter 4 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) $\Theta \not\Rightarrow \varphi$

The first thing to notice is that we must leave the binary format of premises 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 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 $p, q \mid r \not\Rightarrow q$). 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.

2.7.3 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.

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.

2.7.4 Minimal and Preferential Abduction

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 transivity:

- 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 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 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 artesan is willing to represent in a logic. In any case, this format is beyond what we will study in this thesis. It would be a good source, though, for the 'heterogeneous inference' that is coming to the fore these days ([BR97]).

2.7.5 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:

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

We have also investigated in greater detail what happens to these rules when we add our further conditions of 'consistency' and 'explanation' (cf. [Ali94] for a reasoned table of outcomes.) 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 premises, 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 premises 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 thesis, 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 chapter 3.

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 1). 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 premise 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.

2.8 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.

2.8.1 Completeness

The usual completeness theorems have the following form:

$$\Theta \models \varphi \qquad \text{iff} \qquad \Theta \vdash \varphi$$

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

$$\Theta, \alpha \models \varphi \qquad \text{iff} \qquad \Theta, \alpha \vdash \varphi$$

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 relevant 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 succesful (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:

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

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

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

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

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

- Rules For Disjunction:

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

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

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

- Rules for Negation:

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

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

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 $p \wedge \neg p \models p \wedge \neg p$.

Digression: A general semantic view of abductive 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:

$$\Theta \mid \bigvee_i \alpha_i \Rightarrow \varphi \qquad \text{iff} \qquad \text{for some i, } \Theta \mid \alpha_i \Rightarrow \varphi.$$

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 pointwise one, as any set A is the union of all singletons $\{a\}$ with $a \in$ A.

2.8.2 Complexity

Our next question addresses the complexity of different versions of abduction. Nonmonotonic 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_1^0 (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 Φ_1^0. Now, abductive consequence moves further up in this hierarchy.

Proposition 4 *Consistent Abduction is Δ_2^0-complete.*

Proof. The statement that "Θ, α is consistent" is Π_1^0, 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 $\Rightarrow B$. $\dashv$

By similar arguments we can show that the earlier weak explanatory abduction is also 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. We conjecture that consistent abduction will be Δ_2-complete, this time, in the Polynomial Hierarchy.

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 5 *All our notions of abduction 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 is found in [EG95].

2.8.3 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 dissertation.

2.9 Discussion and Conclusions

Studying abduction as a kind of logical inference has provided much more detail to the broad schema in the previous chapter. 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 notions of consequence (classical, preferential, statistical). The resulting abductive 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.

Here is what we consider the main outcomes of our analysis. We can see abductive 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. [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.

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 real logicians. Also, we do not claim that this analysis exhausts all essential features of abduction, as discussed in chapter 1. 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 abductions 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 4. 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.

2.10 Further Questions

We finally indicate a few further issues that we have considered in our work, but that did not make it into our main exposition. These take the form of open questions, or merely promising directions.

(1) To provide complete structural characterizations of all abductive styles put forward in this chapter. In particular, to characterize full explanatory abduction, which accumulates all our constraints.

(2) To relate our deviant structural rules to specific search strategies for abduction.

(3) To provide complete calculi for our styles of abduction with additional logical connectives.

(4) To provide another analysis of abduction, not via the 'amendment strategy' of this chapter, but via some new semantic primitives – as is done in intuitionistic or relevant logic. (This might also decrease complexity.)

(5) To explore purely proof-theoretic approaches, where abduction serves to 'fill gaps' in given arguments. The relevant parameters will then also include some argument, and not just (sets of) assertions.

(6) In line with the previous suggestion, to give a full exploration of abduction in the setting of Toulmin's argumentation theory.

(7) To analyze abductions where the explanation involves changing vocabulary. (Even Bolzano already considered such inferences, which may be related to interpolation theorems in standard logic.) More ambitiously, the 'anomaly' version of this would lead to logical theories of concept revision.

2.11 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 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. 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 the 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[6]. 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).

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

[6]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).

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 implies that it is invalid, cf. chapter 4). 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 [7]. 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. For example, in his setting consistency is enforced by adding the condition of reflexivity ($\gamma \prec \gamma$) to the rules which require it (recall that reflexivity is only allowed for consistent formulae), a style reflecting the methodology of [KLM90].

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.

[7] 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.

Chapter 3

Abduction as Computation

3.1 Introduction

Our logical analysis of abduction in the previous chapter is in a sense, purely structural. It was possible to state how abductive logic behaves, but not how abductions are generated. In this chapter we turn to the question of abduction as a computational process. There are several frameworks for computing abductions; two of which are logic programming and semantic tableaux. The former is a popular one, and it has opened a whole field of abductive logic programming [KKT95]. The latter has also been proposed for handling abduction [MP93], and it is our preference here. Semantic tableaux are a well-motivated standard logical framework. But over these structures, different search strategies can compute several versions of abduction with the non-standard behaviour that we observed in the preceding chapter. Moreover, we can naturally compute various kinds of abducibles: atoms, conjunctions or even conditionals. This goes beyond the framework of abductive logic programming, in which abducibles are atoms from a special set of abducibles.

This chapter is naturally divided into three parts. We first propose abduction as a process of tableau expansion, with each abductive version corresponding to some appropriate 'tableau extension' for the background theory. In the second part, we put forward an algorithm to compute these different abductive versions. In particular, explanations with complex forms are constructed from simpler ones. This allows us to identify cases without consistent atomic explanations whatsoever. It also suggests that in practical implementations of abduction, one can implement our views on different abductive *outcomes* in chapter 1. The third part discusses various logical aspects of tableau abduction, including further semantic analysis, validity of structural rules as studied in chapter 2, plus soundness and completeness of our algorithms.

Generally speaking, this chapter shows how to implement abduction, how to provide procedural counterparts to the abductive versions described in chapter 2. There are still further uses, though which go beyond our analysis so far. Abduction as revision can also be implemented in semantic tableaux. Chapter 4 will demonstrate this, when elaborating a connection with theories of belief change in AI. A detailed description of our algorithms, as well as an implementation in Prolog code, follow in Appendix A.

3.2 Procedural Abduction

3.2.1 Computational Perspectives

There are several options for treating abduction from a procedural perspective. One is standard proof analysis, as in logical proof theory (cf. [Tro96]) or in special logical systems that depend very much on proof-theoretic motivations, such as relevant logic, or linear logic. Proof search via the available rules would then be the driving force for finding abducibles. Another approach would program purely computational algorithms to produce the various types of abduction that we want. An intermediate possibility is logic programming, which combines proof theory with an algorithmic flavor. The latter is more in line with our view of abductive logic as inference plus a control strategy (cf. chapter 1). Although we will eventually choose yet a different route toward the latter end, we do sketch a few features of this practically important approach.

3.2.2 Abducing in Logic Programming

Computation of abductions in logic programming can be formulated as the following process. We wish to produce literals $\alpha_1, \ldots, \alpha_n$ which, when added to the current program P as new facts, make an earlier failed goal φ (the 'surprising fact') succeed after all via Prolog computation $\Rightarrow_p$:

α is an abductive explanation for query φ given program P

if $P \Rightarrow_p \neg\varphi$, while $(\alpha \leftarrow), P \Rightarrow_p \varphi$

Notice that we insert the abducibles as facts into the program here - as an aid. It is a feature of the Prolog proof search mechanism, however, that other positions might give different derivational effects. In this chapter, we merely state these, and other features of resolution-style theorem proving without further explanation, as our main concerns lie elsewhere.

Two Abductive Computations

Abductions are produced via the PROLOG resolution mechanism and then checked against a set of 'potential abducibles'. But as we just noted, there are several ways to characterize an abductive computation, and several ways to add a fact to a Prolog program. An approach which distinguishes between two basic abductive procedures is found in [Sti91], who defines *most specific abduction* (MSA) and *least specific abduction* (LSA). These differ as follows. Via MSA only pure literals are produced as abductions, and via LSA those that are not[1]. The following example illustrates these two procedures (it is a combination of our earlier common sense rain examples):

Program P**:** $r \leftarrow c$, $w \leftarrow r$, $w \leftarrow s$

Query q: w

MSA: c, s

LSA: r

[1]A literal is a 'pure literal' for program P if it cannot be resolved via any clause in the program. A 'nonpure literal' on the other hand, is one which only occurs in the body of a clause but never as a head.

This distinction is useful when we want to identify those abductions which are 'final causes' (MSA) from 'indirect causes' which may be explained by something else (LSA).

Structural Rules

This type of framework also lends itself to a study of logical structural rules like in chapter 2. This time, non-standard effects may reflect computational peculiarities of our proof search procedure. (Cf. [Min90, Kal95, vBe92] for more on this general phenomenon.) As for Monotonicity, we have already shown (cf. chapter 2) that right- but not left-insertion of new clauses in a program is valid for Prolog computation. Thus, adding an arbitrary formula at the beginning of a program may invalidate earlier programs. (With inserting atoms, we can be more liberal: but cf. [Kal95] for pitfalls even there.) For another important structural rule, consider Reflexivity. It is valid in the following form:

Reflexivity: $\alpha \leftarrow, P \Rightarrow_p \alpha$

but invalid in the form

Reflexivity: $P, \alpha \leftarrow, \Rightarrow_p \alpha$

Moreover, these outcomes reflect the downward computation rule of Prolog. Other algorithms can have different structural rules[2].

3.3 Introduction to Semantic Tableaux

3.3.1 Tableau Construction

The logical framework of semantic tableaux is a refutation method introduced in the 50's independently by Beth [Bet69] and Hintikka [Hin55]. A more modern version

[2]In particular, success or failure of Reflexivity may depend on whether a 'loop clause' $\alpha \rightarrow \alpha$ is present in the program. Also, Reflexivity is valid under LSA, but not via MSA computation, since a formula implied by itself is not a pure literal.

is found in [Smu68] and it is the one presented here. The general idea of semantic tableaux is as follows:

> To test if a formula φ follows from a set of premises Θ, a *tableau tree* for the sentences in $\Theta \cup \{\neg\varphi\}$ is constructed. The tableau itself is a binary tree built from its initial set of sentences by using rules for each of the logical connectives that specify how the tree branches. If the tableau closes, the initial set is unsatisfiable and the entailment $\Theta \models \varphi$ holds. Otherwise, if the resulting tableau has open branches, the formula φ is not a valid consequence of Θ. A tableau closes if every branch contains an atomic formula β and its negation.

The rules for constructing the tableau tree are the following. Double negations are suppressed. True conjunctions add both conjuncts, negated conjunctions branch into two negated conjuncts. True disjunctions branch into two true disjuncts, while negated disjunctions add both negated disjuncts. Implications $(a \rightarrow b)$ are treated as disjunctions $(\neg a \vee b)$.

- Negation

$$\neg\neg X \qquad \longrightarrow \qquad X$$

- Conjunction

$$X \wedge Y \qquad \longrightarrow \qquad \frac{X}{Y}$$

$$\neg(X \wedge Y) \qquad \longrightarrow \qquad \neg X \mid \neg Y$$

- Disjunction

$$X \vee Y \qquad \longrightarrow \qquad X \mid Y$$

$$\neg(X \vee Y) \qquad \longrightarrow \qquad \frac{\neg X}{\neg Y}$$

- Implication

$$X \to Y \qquad \longrightarrow \qquad \neg X \mid Y$$

$$\neg(X \to Y) \qquad \longrightarrow \qquad \frac{X}{\neg Y}$$

These seven rules for tableaux construction reduce to two general types, one 'conjunctive' (α-type) and one 'disjunctive' (β-type). The former for a true conjunction and the latter for a true disjunction suffice if every formula to be incorporated into the tableau is transformed first into a propositional conjunctive or a disjunctive normal form.

Rule A:

$$\alpha \qquad \longrightarrow \qquad \frac{\alpha_1}{\alpha_2}$$

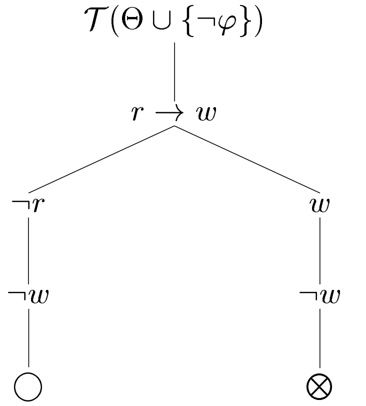
Rule B:

$$\beta \qquad \longrightarrow \qquad \beta_1 \mid \beta_2$$

A Simple Example

To show how this works, we give an extremely simple example. More elaborate tableaux will be found in the course of this chapter.

Let $\Theta = \{r \to w\}$. Set $\varphi = w$. We ask whether $\Theta \models \varphi$. The tableau is as follows. Here, an empty circle $\bigcirc$ indicates that a branch is open, and a crossed circle $\otimes$ that the branch is closed:



The resulting tableau is open, showing that $\Theta \not\models \varphi$. The open branch indicates a 'counterexample', that is, a case in which Θ is true while φ is false (r, w false). More generally, the construction principle is this. A tableau has to be expanded by applying the construction rules until formulas in the nodes have no connectives, and have become literals (atoms or their negations). Moreover, this construction process ensures that each node of the tableau can only carry a subformula of Θ or $\neg\varphi$.

3.3.2 Logical Properties

The tableau method as sketched so far has the following general properties. These can be established by simple analysis of the rules and their motivation, as providing an exhaustive search for a counter-example. In what follows, we concentrate on verifying the top formulas, disregarding the initial motivation of finding counterexamples to consequence problems. This presentation incurs no loss of generality. Given a tableau for a theory ($\mathcal{T}(\Theta)$):

- If $\mathcal{T}(\Theta)$ has open branches, Θ is consistent. Each open branch corresponds to a verifying model.
- If $\mathcal{T}(\Theta)$ has all branches closed, Θ is inconsistent.

Another, more computational feature is that, given some initial verification problem, the order of rule application in a tableau tree does not affect the result. The structure of the tree may be different, but the outcome as to consistency is the same. Moreover, returning to the formulation with logical consequence problems, we have that semantic tableaux are a *sound and complete* system:

$$\Theta \models \varphi \qquad \text{iff} \qquad \text{there is a closed tableau for } \Theta \cup \{\neg\varphi\}.$$

Given the workings of the above rules, which decrease complexity, tableaux are a decision method for propositional logic. This is different with predicate logic (not treated here), where quantifier rules may lead to unbounded repetitions. In the latter case, the tableau method is only semi–decidable. (If the initial set of formulas is unsatisfiable, the tableau will close in finitely many steps. But if it is satisfiable, the tableau may become infinite, without terminating, recording an infinite model.) In this chapter, we shall only consider the propositional case.

A more combinatorial observation is that there are two faces of tableaus. When an entailment does not hold, read upside down, open branches are records of counter-examples. When the entailment does hold, read bottom up, a closed tableau is easily reconstructed as a Gentzen sequent calculus proof. This is no accident of the method. In fact, Beth's motivation for inventing tableaux was his desire to find a combination of proof *analysis* and proof *synthesis*, as we already observed in chapter 2.

Tableaux are widely used in logic, and they have many further interesting properties. For a more detailed presentation, the reader is invited to consult [Smu68, Fit90].

For convenience in what follows, we give a quick reference list of some major notions concerning tableaus.

Closed Branch : A branch of a tableau is closed if it contains some formula and its negation.

Atomically Closed Branch : A branch is atomically closed if it is closed by an atomic formula or a negation thereof.

Open branch : A branch of a tableau is open if it is not closed.

Complete branch : A branch B of a tableau is complete if (referring to the earlier-mentioned two main formula types) for every α which occurs in B, both α_1 and α_2 occur in B, and for every β which occurs in B, at least one of β_1, β_2 occurs in B.

Completed Tableau : A tableau $\mathcal{T}$ is completed if every branch of $\mathcal{T}$ is either closed or complete.

Proof of X : A proof of a formula X is a closed tableau for $\neg X$.

Proof of $\Theta \models \varphi$: A proof of $\Theta \models \varphi$ is a closed tableau for $\Theta \cup \{\neg\varphi\}$.

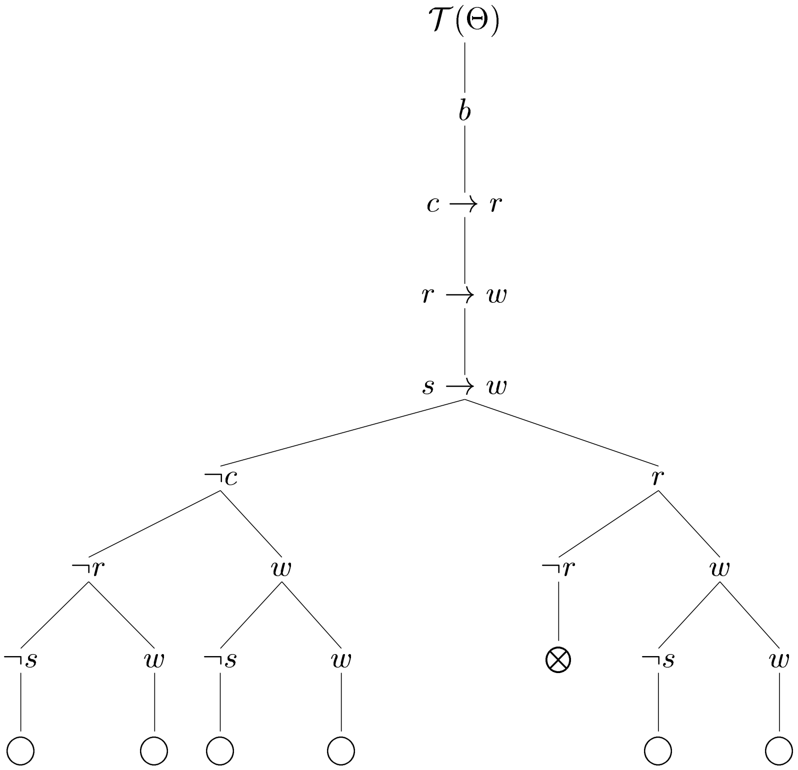
3.4 Abduction with Tableaux

In this section we will show the main idea for performing abduction as a kind of tableau extension. First of all, a tableau itself can represent finite theories. We show this by a somewhat more elaborate example. To simplify the notation from now on, we write $\Theta \cup \neg\varphi$ for $\Theta \cup \{\neg\varphi\}$ (that is, we omit the brackets).

Example

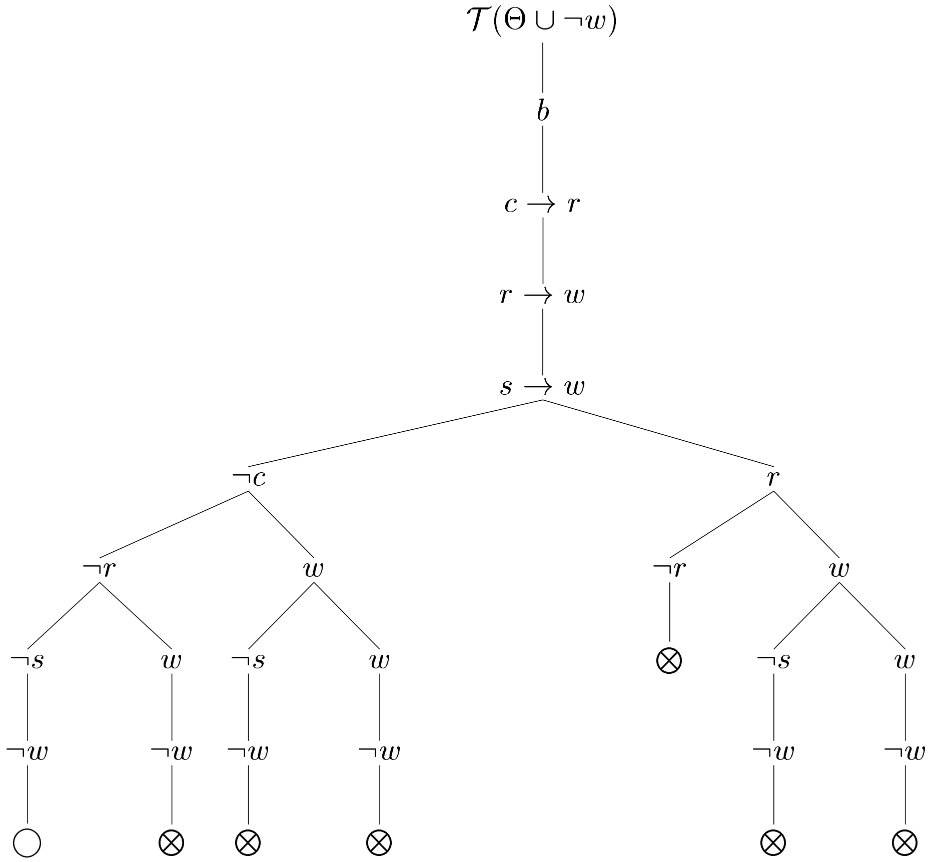
Let $\Theta = \{b, c \rightarrow r, r \rightarrow w, s \rightarrow w\}$, and let $\varphi = \{w\}$.

A tableau for Θ is as follows:



The result is an open tableau. Therefore, the theory is consistent and each open branch corresponds to a verifying model. For example, the second branch (from left to right) indicates that a model for Θ is given by making c, r false and b, w true, so we get two possible models out of this branch (one in which s is true, the other in which it is false). Generally speaking, when constructing the tableau, the possible valuations for the formulas are depicted by the branches (either $\neg c$ or r makes the first split, then for each of these either $\neg r$ or w, and so on).

When formulas are added (thereby extending the tableau), some of these possible models may disappear, as branches start closing. For instance, when $\neg\varphi$ is added (i.e. $\neg w$), the result is the following:



Notice that, although the resulting theory remains consistent, all but one branch has closed. In particular, most models we had before are no longer valid, as w is no longer true. There is still an open branch, indicating there is a model satisfying $\Theta \cup \neg w$ (c, r, s, w false, b true), which indicates that $\Theta \not\models w$.

3.4.1 The Main Ideas

An attractive feature of the tableau method is that when φ is not a valid consequence of Θ, we get all cases in which the consequence fails graphically represented by the open branches (as shown above, the latter may be viewed as descriptions of models for $\Theta \cup \neg\varphi$.)

This fact suggests that if these *counterexamples* were 'corrected by amending the theory', through adding more premises, we could perhaps make φ a valid consequence of some (minimally) extended theory Θ'. This is indeed the whole issue of abduction. Accordingly, abduction may be formulated in this framework as a process of *expansion*, extending a tableau with suitable formulas that close the open branches.

In our example above, the remaining open branch had the following relevant (literal) part:

$$\begin{array}{c} b \\ | \\ \neg c \\ | \\ \neg r \\ | \\ \neg s \\ | \\ \neg w \\ | \\ \bigcirc \end{array}$$

The following are (some) formulas whose addition to the tableau would close this branch (and hence, the whole tableau):

$$\{\neg b, c, r, s, w, c \wedge r, r \wedge w, s \wedge w, s \wedge \neg w, c \vee w\}$$

Note that several forms of statement may count here as abductions. In particular, those in disjunctive form (e.g. $c \vee w$) create two branches, which then both close. (We will consider these various cases in detail later on.)

3.5 Generating Abductions in Tableaux

In principle, we can compute abductions for all our earlier abductive versions (cf. chapter 2). A direct way of doing so is as follows:

> First compute abductions according to the plain version and then eliminate all those which do not comply with the various additional requirements.

This strategy first translates our abductive formulations to the setting of semantic tableaux as follows:

> Given Θ (a set of formulae) and φ (a sentence), α is an abductive explanation if:
>
> **Plain** :
>
> $\mathcal{T}((\Theta \cup \neg\varphi) \cup \alpha)$ is closed. $\quad (\Theta, \alpha \models \varphi)$.
>
> **Consistent** : Plain Abduction +
>
> $\mathcal{T}(\Theta \cup \alpha)$ is open $\quad (\Theta \not\models \neg\alpha)$
>
> **Explanatory** : Plain Abduction +
>
> (i) $\mathcal{T}(\Theta \cup \neg\varphi)$ is open $\quad (\Theta \not\models \varphi)$
>
> (ii) $\mathcal{T}(\alpha \cup \neg\varphi)$ is open $\quad (\alpha \not\models \varphi)$

In addition to the 'abductive conditions' we must state constraints over our search space for abducibles, as the set of formulas fulfilling any of the above conditions is in principle infinite. Therefore, we impose restrictions on the vocabulary as well as on the form of the abduced formulas:

- Restriction on Vocabulary

  α is in the vocabulary of the theory and the observation:

  $\alpha \in \mathrm{Voc}(\Theta \cup \{\varphi\})$.

- Restriction on Form

  The syntactic form of α is either a literal, a conjunction of literals (without repeated conjuncts), or a disjunction of literals (without repeated disjuncts).

Once it is clear what our search space for abducibles is, we continue with our discussion. Note that while computation of plain abductions involves only closed branches, the other versions use inspection of both closed and open branches. For example, an algorithm computing consistent abductions would proceed in the following two steps:

- *Generating Consistent Abductions* (First Version)

  1. Generate all plain abductions, being those formulas α such that $\mathcal{T}((\Theta \cup \neg\varphi) \cup \alpha)$ is closed.
  2. Take out all those α for which $\mathcal{T}(\Theta \cup \alpha)$ is closed.

In particular, an algorithm producing consistent abductions along these lines must produce all explanations that are inconsistent with Θ. This means many ways of closing $\mathcal{T}(\Theta)$, which will then have to be removed in Step 2. This is of course wasteful. Even worse, when there are no consistent explanations (besides the trivial one), so that we would want to give up, our procedure still produces the inconsistent ones. The same point holds for our other versions of abduction.

Of course, there is a preference for procedures that generate abductions in a reasonably efficient way. We will show how to devise these, making use of the representation structure of tableaux, in a way which avoids the production of inconsistent formulae. Here is our idea.

- *Generating Consistent Abductions* (Second Version)

  1. Generate all formulas α which close some (but not all) open branches of $\mathcal{T}(\Theta)$.
  2. Check which of the formulas α produced are such that $\mathcal{T}((\Theta \cup \neg\varphi) \cup \alpha)$ is closed.

That is, first produce formulas which extend the tableau for the background theory in a consistent way, and then check which of these are abductive explanations. In our example above the difference between the two algorithmic versions is as follows (taking into account only the atomic formulas produced). Version 1 produces a formula $(\neg b)$ which is removed for being inconsistent, and version 2 produces a consistent formula $(\neg w)$ which is removed for not being explanatory. As we will show later, the consistent formulae produced by the second procedure are not necessarily wasteful. They might be 'partial explanations' (an ingredient for explanations in conjunctive form), or part of explanations in disjunctive form.

In other words, consistent abductions are those formulas which *"if they had been in the theory before, they would have closed those branches which remain open after* $\neg\varphi$ *is incorporated into the tableau.*

In order to implement our second algorithm, we need to introduce some further distinctions into the tableau framework. More precisely, we need different ways in which a tableau may be extended by a formula. In the next section we define such extensions.

3.6 Tableaux Extensions and Closures

3.6.1 Tableaux extensions: informal explanation

A tableau is extended with a formula via the usual expansion rules (explained in section 3). An extension may modify a tableau in several ways. These depend both on the form of the formula to be added and on the other formulas in the theory represented in the original tableau. If an atomic formula is added, the extended tableau is just like the original with this formula appended at the bottom of its open branches. If the formula has a more complex form, the extended tableau may look quite different (e.g., disjunctions cause every open branch to split into two). In total, however, when expanding a tableau with a formula, the effect on the open branches can only be of three types. Either (i) the added formula closes no open branch or (ii) it closes all open branches, or (iii) it may close some open branches while leaving others open. In order to compute consistent and explanatory abductions, we need to clearly distinguish these three ways of extending a tableau. We label them as 'open', 'closed', and 'semi-closed' extensions, respectively. In what follows we define these notions more precisely.

3.6.2 Formal Definitions

A propositional language is assumed with the usual connectives, whose formulas are of three types: literals (atoms or their negations), α-type (conjunctive form), or β-type (disjunctive form) (cf. section 3).

Completed Tableaux

A completed tableau for a theory ($\mathcal{T}(\Theta)$) is represented as the union of its branches. Each set of branches is the set of formulas which label that branch.

$$\mathcal{T}(\Theta) = \Gamma_1 \cup \ldots \cup \Gamma_k \qquad \text{where each } \Gamma_i \text{ may be open or closed.}$$

Our treatment of tableaux will be always on completed tableau (cf. section 3), so we just refer to them as tableaux from now on.

The Extension Operation

Given $\mathcal{T}(\Theta)$ the addition of a formula γ to each of its branches Γ is defined by the following + operation:

- Γ closed: $\Gamma + \gamma = \Gamma$
- Γ is a completed open branch:

  Case 1 γ is a literal

  $\Gamma + \gamma = \Gamma \cup \{\gamma\}$

  Case 2 γ is an α-type ($\gamma = \alpha_1 \wedge \alpha_2$).

  $\Gamma + \gamma = ((\Gamma \cup \{\gamma\}) + \alpha_1) + \alpha_2$

  Case 3 γ is a β-type ($\gamma = \beta_1 \vee \beta_2$).

  $\Gamma + \gamma = \{(\Gamma \cup \{\gamma\}) + \beta_1), ((\Gamma \cup \{\gamma\}) + \beta_2\}$

That is, the addition of a formula γ to a branch is either Γ itself when it is closed or it is the union of its resulting branches. The operation $+$ is defined over branches, but it easily generalizes to tableaux as follows:

- Tableau Extension:

  $\mathcal{T}(\Theta) + \{\gamma\} =_{def} \bigcup\{\Gamma + \gamma \mid \Gamma \in \mathcal{T}(\Theta)\}$

Our notation allows also for embeddings $((\Theta + \gamma) + \beta)$. Note that operation $+$ is just another way of expressing the usual tableau expansion rules (cf. section 3). Therefore, each tableau may be viewed as the result of a suitable series of $+$ extension steps, starting from the empty tableau.

Branch Extension Types

Given an open branch Γ and a formula γ, we have the following possibilities to extend it:

- Open Extension:

  $\Gamma + \gamma = \delta_1 \cup \ldots \cup \delta_n$ is open if each δ_i is open.

- Closed Extension:

  $\Gamma + \gamma = \delta_1 \cup \ldots \cup \delta_n$ is closed iff each δ_i is closed.

- Semi-Closed Extension:

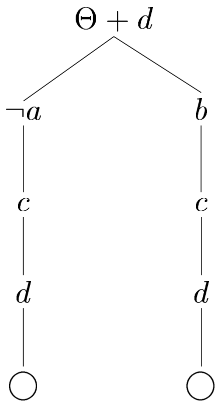
  $\Gamma + \gamma = \delta_1 \cup \ldots \cup \delta_n$ is semi-closed iff at least one δ_i is open and at least one δ_j is closed.

Extensions can also be defined over whole tableaux by generalizing the above definitions. A few examples will illustrate the different situations that may occur.

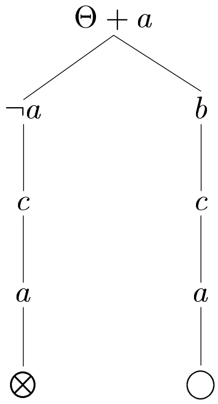
Examples

Let $\Theta = \{\neg a \vee b, c\}$.

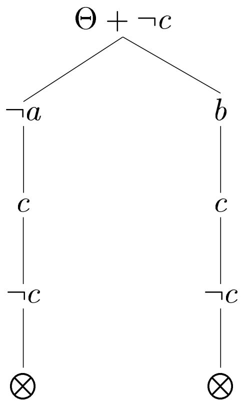
- Open Extension: $\Theta + d$ (d closes no branch).



- Semi-Closed Extension: $\Theta + a$ (a closes only one branch).



- Closed Extension: $\Theta + \neg c$ ($\neg c$ closes all branches)



Finally, to recapitulate an earlier point, these types of extension are related to consistency in the following way:

- Consistent Extension:

  If $\Theta + \gamma$ is open or semi-closed, then $\Theta + \gamma$ is a consistent extension.

- Inconsistent Extension:

  If $\Theta + \gamma$ is closed, then $\Theta + \gamma$ is an inconsistent extension.

3.6.3 Branch and Tableau Closures

As we have stated in section 4, given a theory Θ and a formula φ, plain abductive explanations are those formulas which close the open branches of $\mathcal{T}(\Theta \cup \neg\varphi)$. Furthermore, we suggested that consistent abductive explanations are amongst those formulas which close some (but not all) open branches of $\mathcal{T}(\Theta)$.

In order to compute both kinds, we need to define 'total' and 'partial closures' of a tableau. The first is the set of all literals which close every open branch of the tableau, the second of those literals which close some but not all open branches. For technical convenience, we define total and partial closures for both branches and tableaux. We also need an auxiliary notion. The negation of a literal is either its ordinary negation (if the literal is an atom), or else the underlying atom (if the literal is negative).

Given $\mathcal{T}(\Theta)= \{\Gamma_1, \ldots, \Gamma_n\}$ (here the Γ_i are just the open branches of $\mathcal{T}(\Theta)$):

Branch Total Closure (BTC) :

The set of literals which close an open branch Γ_i:

$$\mathrm{BTC}(\Gamma_i) = \{x \mid \neg x \in \Gamma_i\}, \qquad \text{where } x \text{ ranges over literals.}$$

Tableau Total Closure (TTC) :

The set of those literals which close all branches at once,i.e. the intersection of the BTC's:

$$TTC(\Theta) = \bigcap_{i=1}^{i=n} BTC(\Gamma_i)$$

Branch Partial Closure (BPC) :

The set of those literals which close the branch but do not close all the other open branches:

$$BPC(\Gamma_i) = BTC(\Gamma_i) - TTC(\Theta)$$

Tableau Partial Closure (TPC) :

The set formed by the union of BPC, i.e. all those literals which partially close the tableau:

$$TPC(\Theta) = \bigcup_{i=1}^{i=n} BPC(\Gamma_i)$$

In particular, the definition of BPC may look awkward, as it defines partial closure in terms of branch and tableau total closures. Its motivation lies in a way to compute what we will later call *partial explanations*, being formulas which do close some branches (so they do 'explain') without closing all (so they are 'partial'). We will use the latter to construct explanations in conjunctive form.

Having defined all we need to exploit the framework of semantic tableau for our purposes, we proceed to the construction of abductive explanations.

3.7 Computing Plain Abductions

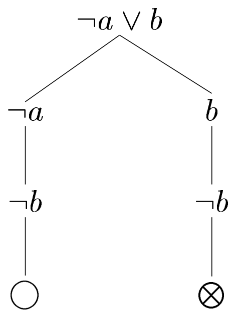
Our strategy for computing plain abduction in semantic tableaux will be as follows. We will be using tableaux as an ordinary consequence test, while being careful about the search space for potential abducibles. The computation is divided into different forms of explanations. Atomic explanations come first, followed by conjunctions of literals, to end with those in disjunctive form. Here we sketch the main ideas for their construction, and give an example for each kind. The detailed algorithms for each case are described in Appendix A to the thesis.

3.7.1 Varieties of Abduction

Atomic Plain Abduction

The idea behind the construction of atomic explanations is very simple. One just computes those atomic formulas which close every open branch of $\mathcal{T}(\Theta \cup \neg\varphi)$, corresponding precisely to its Total Tableaux Closure (TTC($\Theta \cup \neg\varphi$)). Here is an example:

Let $\Theta = \{\neg a \vee b\}$ $\qquad \varphi = b$.



The two possible atomic plain abductions are $\{a, b\}$.

Conjunctive Plain Abduction

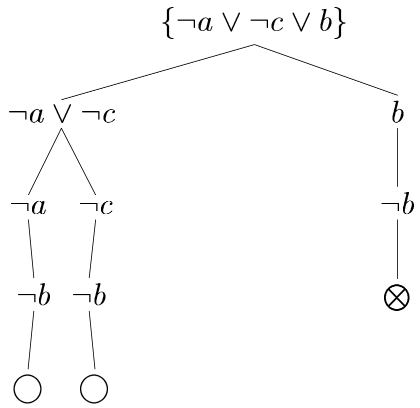
Single atomic explanations may not always exist, or they may not be the only ones of interest. The case of explanations in conjunctive form ($\alpha = \alpha_1 \wedge \ldots \wedge \alpha_n$) is similar to the construction of atomic explanations. We look for literals that close branches, but in this case we want to get the literals that close some but not all of the open branches. These are the conjuncts of a 'conjunctive explanation', and they belong to the tableau partial closure of Θ (i.e., to $\text{TPC}(\Theta \cup \neg\varphi)$). Each of these *partial explanations* make the fact φ 'less surprising' by closing some of the open branches. Together they constitute an abductive explanation.

As a consequence of this characterization, no partial explanation is an atomic explanation. That is, a conjunctive explanation must be a conjunction of partial explanations. The motivation is this. We want to construct explanations which are *non-redundant*, in which every literal does some explaining. Moreover, this condition allows us to bound the production of explanations in our algorithm. We do not want to create what are intuitively 'redundant' combinations. For example, if p and q are abductive explanations, then $p \wedge q$ should not be produced as explanation. Thus we impose the following condition:

> *Non-Redundancy*
>
> Given an abductive explanation α for a theory Θ and a formula φ, α is *non-redundant* if it is either atomic, or no subformula of α (different from φ), is an abductive explanation.

The following example gives an abductive explanation in conjunctive form which is non-redundant:

Let $\Theta = \{\neg a \vee \neg c \vee b\}$, and $\varphi = b$. The corresponding tableau is as follows:



The only atomic explanation is the trivial one $\{b\}$. The conjunctive explanation is $\{a \wedge c\}$ of which neither one is an atomic explanation.

Disjunctive Plain Abductions

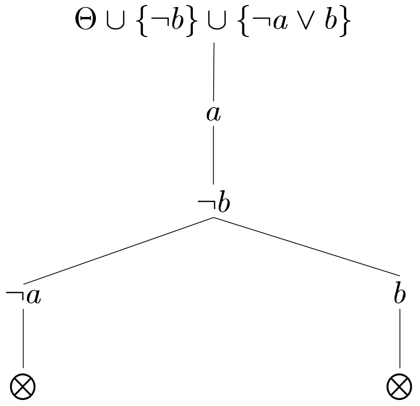
To stay in line with computational practice, we shall sometimes regard abductive explanations in disjunctive form as implications. (This is justified by the propositional equivalence between $\neg\alpha_i \vee \alpha_j$ and $\alpha_i \rightarrow \alpha_j$.) These special explanations close a branch by splitting it first into two. Disjunctive explanations are constructed from atomic and partial explanations. We will not analyze this case in full detail, but provide an example of what happens.

Let $\Theta = \{a\} \qquad \varphi = b.$

The tableau structure for $\mathcal{T}(\Theta \cup \neg b)$ is as follows:

$$\begin{array}{c} \Theta \cup \neg b \\ | \\ a \\ | \\ \neg b \\ | \\ \bigcirc \end{array}$$

Notice first that the possible atomic explanations are $\{\neg a, b\}$ of which the first is inconsistent and the second is the trivial solution. Moreover, there are no 'partial explanations' as there is only one open branch. An explanation in disjunctive form is constructed by combining the atomic explanations: $\{\neg a \vee b\}$. The effect of adding it to the tableau is as follows:



This examples serves as a representation of our example in chapter 1, in which a causal connection is found between certain type of clouds (a) and rain (b), namely that a causes b ($a \rightarrow b$).

3.7.2 Algorithm for Computing Plain Abductions

The general points of our algorithm for computing plain abductions is displayed here (Cf. Appendix A for the more detailed description).

- **Input:**

  . A set of propositional formulas separated by commas representing the theory Θ.

  . A literal formula φ representing the 'fact to be explained'.

  . Preconditions: Θ, φ are such that $\Theta \not\models \varphi$, $\Theta \not\models \neg\varphi$.

- **Output:**

  Produces the set of abductive explanations: $\alpha_1, \ldots \alpha_n$ such that:

  (i) $\mathcal{T}((\Theta \cup \neg\varphi) \cup \alpha_i)$ is closed.

  (ii) α_i complies with the vocabulary and form restrictions (cf. section 5).

- **Procedure:**

  . Calculate $\Theta + \neg\varphi = \{\Gamma_1, \ldots, \Gamma_k\}$

  . Take those Γ_i which are open branches:$\Gamma_1, \ldots, \Gamma_n$

  . **Atomic Plain Explanations**

1. Compute TTC($\Gamma_1, \ldots, \Gamma_n$)= $\{\gamma_1, \ldots, \gamma_m\}$.
2. $\{\gamma_1, \ldots, \gamma_m\}$ is the set of atomic plain abductions.

. **Conjunctive Plain Explanations**

1. For each open branch Γ_i, construct its partial closure: BPC(Γ_i).
2. Check if all branches Γ_i have a partial closure, for otherwise there cannot be a conjunctive solution (in which case, goto END).
3. Each BPC(Γ_i) contains those literals which partially close the tableau. Conjunctive explanations are constructed by taking one literal of each BPC(Γ_i) and making their conjunction. A typical solution is a formula β as follows:
   $a_1 \wedge b_1 \wedge \ldots \wedge z_1 \qquad (a_1 \in BPC(\Gamma_1), b_1 \in BPC(\Gamma_2), \ldots, z_1 \in BPC(\Gamma_n))$
4. Each β conjunctive solution is reduced (there may be repeated literals). The set of solutions in conjunctive form is $\beta_1, \ldots, \beta_l$.
5. END.

. **Disjunctive Plain Explanations**

1. Construct disjunctive explanations by combining atomic explanations amongst themselves, conjunctive explanations amongst themselves, conjunctive with atomic, and each of atomic and conjunctive with φ. We just show two of these constructions:
2. Generate pairs from set of atomic explanations, and construct their disjunctions $(\gamma_i \vee \gamma_j)$.
3. For each atomic explanation construct the disjunction with φ as follows: $(\gamma_i \vee \varphi)$.
4. The result of all combinations above is the set of explanations in disjunctive form.
5. END.

3.8 Consistent Abductive Explanations

The issue now is to compute abductive explanations with the additional requirements of being consistent. For this purpose we will follow the same presentation as for plain abductions (atomic, conjunctive and disjunctive), and will give the key points for their construction. Our algorithm follows version 2 of the strategies sketched earlier (cf. section 5). That is, it first constructs those consistent extensions on the original tableau for Θ which do some closing and then checks which of these is in fact an explanation (i.e. closes the tableau for $\Theta \cup \neg\varphi$). This way we avoid the production of any inconsistency whatsoever. It turns out that in the atomic and conjunctive cases explanations are sometimes necessarily inconsistent, therefore we identify these cases and prevent our algorithm from doing anything at all (so that we do not produce formulae which are discarded afterwards).

3.8.1 Atomic Consistent Abductions

When computing plain atomic explanations, we now want to avoid any computation when there are only inconsistent atomic explanations (besides the trivial one). Here is an observation which helps us get one major problem out of the way. Atomic explanations are necessarily inconsistent when $\Theta + \neg\varphi$ is an open extension. So, we can prevent our algorithm from producing anything at all in this case.

> **Fact 1** *Whenever $\Theta + \neg\varphi$ is an open extension, and α a non-trivial atomic abductive explanation (different from φ), it follows that Θ, α is inconsistent.*
>
> *Proof.* Let $\Theta + \neg\varphi$ be an open extension and α an atomic explanation ($\alpha \neq \varphi$). The latter implies that $((\Theta + \neg\varphi) + \alpha)$ is a closed extension. Therefore, $\Theta + \alpha$ must be a closed extension, too, since φ closes no branches. But then, $\Theta + \alpha$ is an inconsistent extension. I.e. Θ, α is inconsistent. $\dashv$

This result cannot be generalized to more complex forms of abducibles. (We will see later that for explanations in disjunctive form, open extensions need not lead to inconsistency.) In case $\Theta + \neg\varphi$ is a semi-closed extension, we have to do real work, however, and follow the strategy sketched above. The key point in the algorithm is this. Instead of building the tableau for $\Theta \cup \varphi$ directly, and working with its open branches, we must start with the open branches of Θ.

Atomic Consistent Explanations

1. Given $\Theta, \neg\varphi$, construct $\mathcal{T}(\Theta)$, and compute its open branches: $\Gamma_1, \ldots, \Gamma_k$.

2. Compute $\Theta + \neg\varphi$. If it is an open extension, then there are no atomic consistent explanations (by fact 1), GOTO END.

3. Else, compute $\mathrm{TPC}(\Gamma_1, \ldots, \Gamma_k) = \{\gamma_1, \ldots, \gamma_n\}$ which gives those literals which partially close $\mathcal{T}(\Theta)$.

4. Next, check which of α_i close the tableau for $\mathcal{T}(\Theta \cup \neg\varphi)$.

   Consistent Atomic Explanations: $\{\gamma_i \mid \mathcal{T}((\Theta + \neg\varphi) + \gamma_i)$ is closed $\}$.

5. END.

3.8.2 Conjunctive Consistent Explanations

For conjunctive explanations, we can also avoid any computation when there are only 'blatant inconsistencies', essentially by the same observation as before.

Fact 2 *Whenever $\Theta + \neg\varphi$ is an open extension, and $\alpha = \alpha_1 \wedge \ldots \wedge \alpha_n$ is a conjunctive abductive explanation, it holds that Θ, α is inconsistent.*

The proof is analogous to that for the atomic case.

The modification for the algorithm in case $\Theta + \neg\varphi$ is semi-closed, works as follows:

1. For each open branch Γ_i of $\mathcal{T}(\Theta)$, construct its partial closure: $\mathrm{BPC}(\Gamma_i)$.

2. Construct conjunctions of the above (as in the plain case) without taking into account those literals closing the branches where φ appears (to ensure all of them are consistent extensions). Label these conjunctive extensions as $\gamma_1, \ldots, \gamma_l$ respectively.

3. Check which of the γ_i above close the tableau for $\mathcal{T}(\Theta \cup \neg\varphi)$.

   Consistent Conjunctive Explanations $= \{\gamma_i \mid \mathcal{T}\ (\Theta + \neg\varphi) + \gamma_i$ is closed$\}$.

4. END.

3.8.3 Disjunctive Consistent Explanations

As for disjunctive explanations, unfortunately, we no longer have the clear cut distinction between open and semi-closed extensions to know when there are only inconsistent explanations. The reason is that for explanations α in disjunctive form, $(\Theta + \neg\varphi)$ open and $((\Theta + \neg\varphi) + \alpha)$ closed does not imply that $\Theta + \alpha$ is closed because α generates two branches.

We will not write here the algorithm to compute disjunctive consistent explanations (found in appendix A), but instead just present the key issue in its construction:

1. Construct disjunctive formulas by combining the atomic and conjunctive consistent ones above. These are all consistent.

2. Check which of the above formulas close the tableau for $\Theta \cup \neg\varphi$.

What this construction suggests is that there are always consistent explanations in disjunctive form, provided that the theory is consistent:

Fact 3 *Given that $\Theta \cup \neg\varphi$ is consistent, there exists an abductive consistent explanation in disjunctive form.*

The key point to prove this fact is that an explanation may be constructed as $\alpha = \neg X \vee \varphi$, for any $X \in \Theta$.

3.9 Explanatory Abduction

As for explanatory abductions, recall these are those formulas α which are only constructed when the theory does not explain the observation already ($\Theta \not\models \varphi$) and that cannot do the explaining by themselves ($\alpha \not\models \varphi$), but do so in combination with the theory ($\Theta, \alpha \models \varphi$).

Given our previous algorithmic constructions, it turns out that the first condition is already 'built-in', since all our procedures start with the assumption that the tableau for $\Theta \cup \neg\varphi$ is open[3]. As for the second condition, its implementation actually amounts to preventing the construction of the trivial solution ($\alpha = \varphi$). Except for this solution, our algorithms never produce an α such that $\alpha \not\models \varphi$, as proved below:

Fact 4 *Given any Θ and φ, our algorithm never produces abductive explanations α with $\alpha \models \varphi$ (except for $\alpha = \varphi$).*

Proof. Recall our situation: $\Theta \not\models \varphi$, $\Theta, \alpha \models \varphi$, while we have made sure Θ and α are consistent. Now, first suppose that α is a literal. If $\alpha \models \varphi$, then $\alpha = \varphi$ which is the trivial solution. Next, suppose that α is a conjunction produced by our algorithm. If $\alpha \models \varphi$, then one of the conjuncts must be φ itself. (The only other possibility is that α is an inconsistent conjunction, but this is ruled out by our consistency test.) But then, our non-redundancy filter would have produced the relevant conjunct by itself, and then rejected it for triviality. Finally, suppose that α is a disjunctive explanation. Given the above conditions tested in our algorithm, we know that Θ is consistent with at least one disjunct α_i. But also, this disjunct by itself will suffice for deriving φ in the presence of Θ, and it will imply φ by itself. Therefore, by our redundancy test, we would have produced this disjunct by itself, rather than the more complex explanation, and we are in one of the previous cases. $\dashv$

Therefore, it is easy to modify any of the above algorithms to handle the computation of explanatory abductions. We just need to avoid the trivial solution, when $\alpha = \varphi$ and this can be done in the module for atomic explanations.

3.10 Quality of Abductions

A question to ask at this point is whether our algorithms produce intuitively good explanations for observed phenomena. One of our examples (cf. disjunctive plain

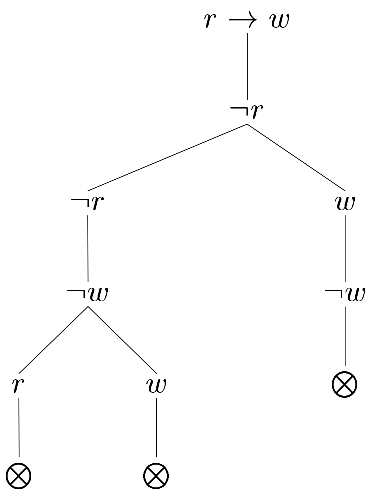
[3]It would have been possible to take out this condition for the earlier versions. However, note that in the case that $\Theta \cup \neg\varphi$ is closed the computation of abductions is trivialized since as the tableau is already closed, any formula counts as a plain explanation and any consistent formula as a consistent abduction.

abductions, section 7.1.3) suggested that abductive explanations for a fact and a theory with no causal knowledge (with no formulas in conditional form) must be in disjunctive form if such a fact is to be explained in a consistent way. Moreover, Fact 3 stated that consistent explanations in disjunctive form are always available, provided that the original theory is consistent. However, producing consistent explanations does not guarantee that these are good or relevant. These further properties may depend upon the nature of the theory itself. If the theory is a bad theory, it will produce bad or weird explanations. The following example illustrates this point.

A Bad Theory

Let $\Theta = \{r \rightarrow w, \neg r\}$ $\qquad \varphi = w \qquad \alpha = \neg r \rightarrow w$.

The tableau structure for $(\Theta \cup \neg\varphi) \cup \alpha$ is depicted as follows:



Interpreted in connection with our rain example, our algorithm will produce the following consistent 'explanation' of why the lawn is wet (w), given that rain causes the lawn to get wet ($r \rightarrow w$) and that it is not raining ($\neg r$). One explanation is that "the absence of rain causes the lawn to get wet" ($\neg r \rightarrow w$). But this explanation seems to trivialize the fact of the lawn being wet, as it seems to be so, regardless of rain!

A better, though more complex, way of explaining this fact would be to conclude that the theory is not rich enough to explain why the lawn is wet, and then look for some external facts to the theory (e.g. sprinklers are on, and they make the lawn wet.) But this would amount to dropping the vocabulary assumption.

Therefore, producing good or bad explanations is not just a business of properly defining the underlying notion of consequence, or of giving an adequate procedure. An inadequate theory like the one above can be the real cause of bad explanations. In other words, what makes a 'good explanation' is not the abducible itself, but the interplay of the abducible with the background theory. Bad theories produce bad explanations. Our algorithm cannot remedy this, only record it.

3.10.1 Discussion

Having a module in each abductive version that first computes only atomic explanations already gives us some account of *minimal explanations* (see chapter 2), when minimality is regarded as simplicity. As for conjunctive explanations, as we have noted before, their construction is one way of computing non-trivial 'partial explanations', which make a fact less surprising by closing some, though not all open branches for its refutation. One might tie up this approach with a more general issue, namely, weaker notions of 'approximative' logical consequence. Finally, explanations in disjunctive form can be constructed in various ways. E.g., one can combine atomic explanations, or form the conjunction of all partial explanations, and then construct a conditional with φ. This reflects our view that abductive explanations are built in a compositional fashion: complex solutions are constructed from simpler ones.

Notice moreover, that we are not constructing all possible formulas which close the open branches, as we have been taking care not to produce what we have called *redundant explanations*. Finally, despite these precautions, as we have noted, bad explanations may slip through when the background theory is inappropriate.

3.11 Further Logical Issues

Our algorithmic tableau analysis suggests a number of further logical issues, which we briefly discuss here.

3.11.1 Rules, Soundness and Completeness

The abductive consequences produced by our tableaux can be viewed as a ternary notion of inference. Its structural properties can be studied in the same way as we did for the more abstract notions of chapter 2. But the earlier structural rules lose some of their point in this algorithmic setting. For instance, it follows from our tableau algorithm that consistent abduction does not allow monotonicity in either its Θ or its α argument. One substitute which we had in chapter 2 was as follows. If $\Theta, \alpha \Rightarrow \varphi$, and $\Theta, \alpha, \beta \Rightarrow \gamma$ (where γ is any conclusion at all), then $\Theta, \alpha, \beta \Rightarrow \varphi$. In our algorithm, we have to make a distinction here. We produce abducibles α, and if we already found α solving $\Theta, \alpha \Rightarrow \varphi$, then the algorithm may not produce stronger abducibles than that. (It might happen, due to the closure patterns of branches in the initial tableau, that we produce one solution implying another, but this does not have to be.) As for strengthening the theory Θ, this might result in an initial tableau with possibly fewer open branches, over which our procedure may then produce weaker abducibles, invalidating the original choice of α cooperating with Θ to derive φ.

More relevant, therefore, is the traditional question whether our algorithms are sound and complete. Again, we have to make sure what these properties mean in this setting. First, Soundness should mean that any combination $(\Theta, \alpha, \varphi)$ which gets out of the algorithm does indeed present a valid case of abduction, as defined in chapter 2. For plain abduction, it is easy to see that we have soundness, as the tableau closure condition guarantees classical consequence (which is all we need). Next, consider consistent abduction. What we need to make sure of now, is also that all abducibles are consistent with the background theory Θ. But this is what happened by our use of 'partial branch closures'. These are sure (by definition) to leave at least one branch for Θ open, and hence they are consistent with it. Finally, the conditions of the 'explanatory' algorithm ensure likewise that the theory does not explain the fact already ($\Theta \not\Rightarrow \varphi$) and that α could not do the job on its own ($\alpha \not\Rightarrow \varphi$).

Next, we consider completeness. Here, we merely make some relevant observations, demonstrating the issues (for our motives, cf. the end of this paragraph). Completeness should mean that any valid abductive consequence should actually be produced by it. This is trickier. Obviously, we can only expect completeness within the restricted language employed by our algorithm. Moreover, the algorithm 'weeds out' irrelevant conjuncts, et cetera, which cuts down outcomes even more. As a more significant source of incompleteness, however, we can look at the case of disjunctive explanations. The implications produced always involve one literal as a consequent. This is not enough for a general abductive conclusion, which might involve more. What we can say, for instance is this. By simple inspection of the algorithm, one can see that every consistent atomic explanation that exists for an abductive problem will be produced by the algorithm. In any case, we feel that completeness is less of an issue in computational approaches to abduction. What comes first is whether a given abductive procedure is natural, and simple. Whether its yield meets some pre-assigned goal is only a secondary concern in this setting.

We can also take another look at issues of soundness and completeness, relatively independently from our axiom. The following analysis of 'closure' on tableaux is inspired by our algorithm - but it produces a less procedural logical view of what is going on.

3.11.2 An Alternative Semantic Analysis

Our strategy for producing abductions in tableaux worked as follows. One starts with a tableau for the background theory Θ (i), then adds the negation $\neg\varphi$ of the new observation φ to its open branches (ii), and one also closes the remaining open branches (iii), subject to certain constraints. In particular (for the explanatory version) one does not allow φ itself as a closure atom as it is regarded as a trivial solution. This operation may be expressed via a kind of 'closure operation':

$$\text{CLOSE } (\Theta + \neg\varphi) - \varphi.$$

We now want to take an independent (in a sense, 'tableau–free') look at the situation, in the most general case, allowing disjunctive explanations. (If Θ is particularly simple, we can make do (as shown before) with atoms, or their conjunctions.) First, we recall that tableaus may be represented as sets of open branches. We may assume that all branches are completed, and hence all relevant information resides in their literals. This leads to the following observation.

Fact 5 *Let Θ be a set of sentences, and let $\mathcal{T}$ be a complete tableau for Θ, with π running over its open branches. Then $\wedge\Theta$ (the conjunction of all sentences in Θ) is equivalent to:*

$$\bigvee_{\pi \text{ open in } \mathcal{T}} \bigwedge_{l \text{ a literal}} l \in \pi.$$

This fact is easy to show. The conjunctions are the total descriptions of each open branch, and the disjunction says that any model for Θ must choose one of them. This amounts to the usual Distributive Normal Form theorem for propositional logic. Now, we can give a description of our CLOSE operation in similar terms. In its most general form, our way of closing an open tableau is really defined by putting:

$$\bigvee_{S \text{ a set of literals}} \bigwedge_{\pi \text{ open in T}} \exists l \in S : l \notin \pi$$

The inner part of this says that the set of (relevant) literals S 'closes every branch'. The disjunction states the weakest combination that will still close the tableau. Now, we have a surprisingly simple connection:

Fact 6 *CLOSE (Θ) is equivalent to $\neg(\wedge\Theta)$!*

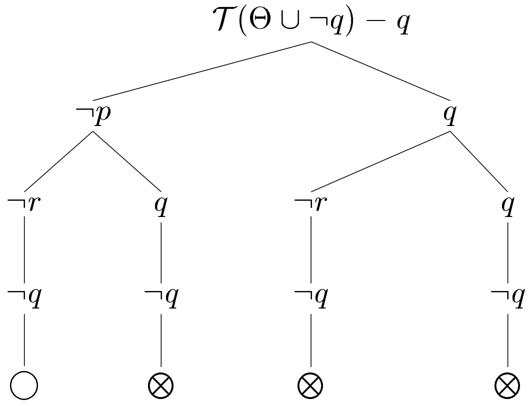
Proof. By Fact 1 plus the propositional De Morgan laws, $\neg(\wedge\Theta)$ is equivalent to $\bigvee_{\pi \text{ open in } \mathcal{T}} \bigwedge_{l \text{ a literal}} l \notin \pi$. But then, a simple argument, involving choices for each open branch, shows that the latter assertion is equivalent to CLOSE (Θ). $\dashv$

In the case of abduction, we proceeded as follows. There is a theory Θ, and a surprising fact q (say), which does not follow from it. The latter shows because we have an open tableau for Θ followed by $\neg q$. We close up its open branches, without using the trivial explanation q. What this involves, as said above, is a modified operation, that we can write as:

CLOSE $(\Theta) - q$

Example

Let $\Theta = \{p \rightarrow q, r \rightarrow q\}$.



The abductions produced are p or r.

Again, we can analyze what the new operation does in more direct terms.

Fact 7 *CLOSE* $(\Theta) - q$ *is equivalent to* $\neg$ *[false/q]* $\wedge \Theta$.

Proof. From its definition, it is easy to see that CLOSE$(\Theta) - q$ is equivalent with [false/q] CLOSE(Θ). But then, we have that

CLOSE $(\Theta \wedge \neg q) - q$ iff (by Fact 2)

[false/q] $\neg(\wedge\Theta \wedge \neg q)$ iff (by propositional logic)

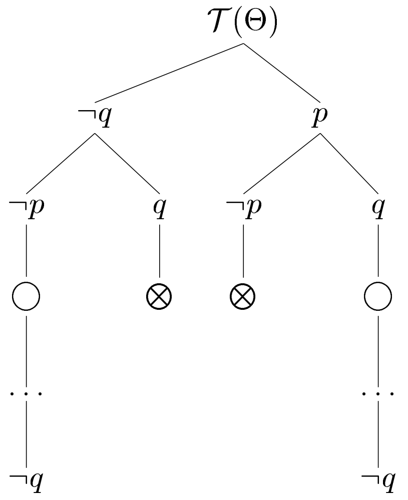
$\neg$[false/q] $\wedge\, \Theta$. $\dashv$

This rule may be checked in the preceding example. Indeed, we have that [false/q] Θ is equivalent to $\neg p \wedge \neg r$, whose negation is equivalent to our earlier outcome $p \vee r$.

This analysis suggests abductive variations that we did not consider before. For instance, we need not forbid all closures involving $\neg q$, but only those which involve $\neg q$ in final position (i.e., negated forms of the 'surprising fact' to be explained). There might be other, harmless occurrences of $\neg q$ on a branch emanating from the background theory Θ itself.

Example

Let $\Theta = \{q \rightarrow p, p \rightarrow q\}$.



In this case, our earlier strategy would merely produce an outcome p - as can be checked by our false-computation rule. The new strategy, however, would compute an abduction p or q, which may be just as reasonable.

This analysis does not extend to complex conclusions. We leave its possible extensions open here.

3.11.3 Tableaux and Resolution

In spite of the logical equivalence between the methods of tableau and resolution [Fit90, Gal92], in actual implementations the generation of abductions turns out to be very different. The method of resolution used in logic programming does not handle negation explicitly in the language, and this fact restricts the kind of abductions to be produced. In addition, in logic programming only atomic formulas are produced as abductions since they are identified as those literals which make the computation fail. In semantic tableaux, on the other hand, it is quite natural to generate abductive formulas in conjunctive or disjunctive form as we have shown.

As for similarities between these two methods as applied to abduction, both frameworks have one of the explanatory conditions ($\Theta \not\Rightarrow \varphi$) already built in. In logic programming the abductive mechanism is put to work when a query fails, and in semantic tableaux abduction is triggered when the tableau for $\Theta \cup \neg\varphi$ is open.

3.12 Discussion and Conclusions

Exploring abduction as a form of computation gave us further insight into this phenomenon. Our concrete algorithms implement some earlier points from chapter 2, which did not quite fit the abstract structural framework. Abductions come in different degrees (atomic, conjunctive, disjunctive-conditional), and each abductive condition corresponds to new procedural complexity. In practice, though, it turned out easy to modify the algorithms accordingly. Indeed these findings reflect an intuitive feature of explanation. While it is sometimes difficult to describe what an explanation is in general, it may be easier to construct a set of explanations for a particular problem.

As for the computational framework, semantic tableaux are a natural vehicle for implementing abduction. They allow for a clear formulation of what counts as an abductive explanation, while being flexible and suggestive as to possible modifications and extensions. Derivability and consistency, the ingredients of consistent and explanatory abduction, are indeed a natural blend in tableaux, because we can manipulate open and closed branches with equal ease. Hence it is very easy to check if the consistency of a theory is preserved when adding a formula. (By the same token, this type of conditions on abduction appears rather natural in this light.)

Even so, our actual algorithms were more than a straight transcription of the logical formulations in chapter 2 (which might be very inefficient). Our computational strategy provided an algorithm which produces consistent formulas, selecting those which count as explanations, and this procedure turns out to be more efficient than the other way around. Nevertheless, abduction in tableaux has no unique form, as we showed by some alternatives. A final insight emerging from a procedural view of abduction is the importance of the background theory when computing explanations. Bad theories will produce bad explanations. Sophisticated computation cannot improve that.

Our tableau approach also has clear limitations. It is hard to treat notions of abduction in which $\Rightarrow$ is some non-standard consequence. In particular, with an underlying statistical inference, it is unclear how probabilistic entailments should be represented. Our computation of abductions relies on tableaux being open or closed, which represent only the two extremes of probable inference. We do have one speculation, though. The computation of what we called 'partial explanations' (which close some but not all open branches) might provide a notion of partial entailment in which explanations only make a fact less surprising, without explaining it in full. (Cf. [Tij97] for other approaches to 'approximative deduction' in abductive diagnostic settings.) As for other possible uses of the tableau framework, the case of abduction as revision was not addressed here. In the following chapter, we shall see that we do get further mileage in that direction, too.

3.13 Further Questions

While working on the issues discussed in this chapter, the following questions emerged, which we list here for further research.

- Fine-structure of tableaux. Our algorithms raise various 'side issues'. For instance, consider tableau expansion. When does a new incoming statement A close no branches of $\mathcal{T}(\Theta)$ at all? Such syntactic observations may have interesting semantic counterparts. The preceding is equivalent to: "$\mathcal{T}$ implies the existential closure of A w.r.t. those proposition letters that do not occur in $\mathcal{T}$". More generally, what would be a general theory of tableaux as a concrete system of 'constructive update logic' (in the sense of [vBe96a])?

- Proof theory revisited. Develop a complete proof theory of abduction, which stands to our use of semantic tableaux as Gentzen sequent calculus stands to ordinary tableaus. (The dual sequent presentation of [MP93] may be suggestive here, but our algorithms are different.)
- From propositional to predicate logic. Extend our algorithm for tableau abduction to first-order predicate logic with quantifiers.
- Structural rules. One would like to have a characterization of the valid structural rules for our abductive algorithms. Do these have additional features, not encountered in chapter 2? At a tangent (cf. [Kal95] for a similar theme in Prolog computation), could one correlate deviant structural rules with new procedures for cut elimination?

3.14 Related Work

The framework of semantic tableaux has recently been used beyond its traditional logical purposes, especially in computationally oriented approaches. One example is found in [Nie96], implementing 'circumscription'. In connection with abduction, semantic tableaux are used in [Ger95] to model natural language presuppositions (cf. chapter 1, abduction in linguistics). A better–known approach is found in [MP93], a source of inspiration for our work, which we proceed to briefly describe and compare. The following is a mere sketch, which cannot do full justice to all aspects of their proposal.

Mayer and Pirri's Tableau Abduction

Mayer and Pirri's article presents a model for computing 'minimal abduction' in propositional and first–order logic. For the propositional case, they propose two characterizations. The first corresponds to the generation of all consistent and minimal explanations (where minimality means 'logically weakest'; cf. chapter 2). The second generates a single minimal and consistent explanation by a non-deterministic algorithm. The first–order case constructs abductions by reversed skolemization, making use of unification and what they call 'dynamic herbrandization' of formulae. To give an idea of their procedures for generating explanations, we merely note their main steps: (1) construction of 'minimal closing sets', (2) construction of abductive solutions as literals which close all branches of those sets, (3) elimination of inconsistent solutions. The resulting systems are presented in two fashions, once as semantic tableaux, and once as sequent calculi for abductive reasoning. There is an elegant treatment of the parallelism between these two. Moreover, a predicate–logical extension is given, which is probably the first significant treatment of first–order abduction which goes beyond the usual resolution framework. (In subsequent work, the authors have been able to extend this approach to modal logic, and default reasoning.) A final interesting point of the Mayer and Pirri presentation is that it shows very well the contrast between computing propositional and first–order abduction. While the former is easy to compute, even considering minimality, the latter is inherently undecidable.

Comparison to our work

Our work has been inspired by [MP93]. But it has gone in different directions, both concerning strategy and output. (i) Mayer and Pirri compute explanations in line with version 1 of the general strategies that we sketched earlier. That is, they calculate all closures of the relevant tableau to later eliminate the inconsistent cases. We do the opposite, following version 2. That is, we first compute consistent formulas which close at least one branch of the original tableau, and then check which of these are explanations. Our reasons for this had to do with greater computational efficiency. (ii) As for the type of explanations produced, Mayer and Pirri's propositional algorithms basically produce minimal atomic explanations or nothing at all, while our approach provides explanations in conjunctive and disjunctive form as well. (iii) Mayer's and Pirri's approach stays closer to the classical tableau framework, while ours gives it several new twists. We propose several new distinctions and extensions, e.g. for the purpose of identifying when there are no consistent explanations at all. (iv) Eventually, we go even further (cf. chapter 4) and propose semantic tableaux as a vehicle for revision, which requires a new contraction algorithm.

Chapter 4

Scientific Explanation and Epistemic Change

4.1 Introduction

In the preceding chapters, notions from the philosophy of science and artificial intelligence have been a source of inspiration. In this chapter, we aim at explicit comparison with these traditions.

In the philosophy of science, we confront our logical account of chapter 2 with the notion of scientific explanation, as proposed by Hempel in two of his models of scientific inference: deductive-nomological and inductive-statistical [Hem65]. We show that both can be viewed as forms of abduction, the former with deductive underlying inference, the latter with statistical inference. As for artificial intelligence, we confront our computational account of chapter 3 with the notion of epistemic change, as modeled by Gärdenfors [Gär88].

In this confrontation, we hope to learn something about the scope and limitations of our analysis so far. We find a number of analogies, as well as new challenges. The notion of statistical inference gives us an opportunity to expand the logical analysis at a point left open in chapter 2, and the dynamics of belief change allows us to extend our treatment of tableaux to the case of revision. We will also encounter further natural desiderata, however, which our current analysis cannot handle.

Our selection of topics in these two fields is by no means complete. Many other connections exist relating our proposal to philosophy of science and knowledge representation in AI. Some of these concerns (such as cognitive processing of natural language) involve abductive traditions beyond the scope of our analysis. (We have already identified some of these in chapter 1.) Nevertheless, the connections that we do develop have a clear intention. We view all three fields as naturally related components in cognitive science, and we hope to show that abduction is one common theme making for cohesion.

4.2 Scientific Explanation as Abduction

At a general level, our discussion in chapter 1 already showed that scientific reasoning can be analyzed as an abductive process. This reasoning comes in different kinds, reflecting (amongst others) various patterns of discovery, with different 'triggers'. A discovery may be made to explain a novel phenomenon which is consistent with our current theory, but it may also point at an anomaly between the theory and the phenomenon observed. Moreover, the results of scientific reasoning vary in their *degree* of novelty and complexity. Some discoveries are simple empirical generalizations from observed phenomena, others are complex scientific theories introducing sweeping new notions. We shall concentrate on rather 'local' scientific explanations, which can be taken to be some sort of logical arguments: abductive ones, in our view. (We are aware of the existence of divergent opinions on this: cf. chapter 1). That scientific inference can be viewed as abductive should not be surprising. Both Peirce and Bolzano were inspired in their logical systems by the way reasoning is done in science. Indeed, Peirce's abductive formulations may be regarded as precursors of Hempel's notion of explanation, as will become clear shortly.

At the center of our discussion on the logic of explanation lies the proposal by Hempel and Oppenheimer [HO48, Hem65]. Their aim was to model explanations of empirical 'why-questions'. For this purpose they distinguished several kinds of explanation, based on the logical relationship between the explanans and explanandum (deductive or inductive), as well as on the form of the explanandum (singular events or general regularities). These two distinctions generate four models altogether: two deductive-nomological ones (D-N), and two statistical ones (Inductive-Statistical (I-S), and Deductive-Statistical (D-S)).

We analyze the two models for singular events, and present them as forms of abduction, obeying certain structural rules.

4.2.1 The Deductive-Nomological Model

The general schema of the D-N model is the following:

$$\begin{array}{l} L_1, \ldots, L_m \\ \underline{C_1, \ldots, C_n \quad} \\ E \end{array}$$

$L_1, \ldots, L_m$ are general laws which constitute a scientific theory T, and together with suitable antecedent conditions $C_1, \ldots, C_n$ constitute a *potential explanation* $\langle T, C \rangle$ for some observed phenomenon E. The relationship between explanandum and explananda is deductive, signaled by the horizontal line in the schema. Additional conditions are then imposed on the explananda:

$\langle T, C \rangle$ is a *potential explanation* of E iff

- T is essentially general and C is singular.
- E is derivable from T and C jointly, but not from C alone.

The first condition requires T to be a 'general' theory (having at least one universally quantified formula). A 'singular' sentence C has no quantifiers or variables, but just closed atoms and Boolean connectives. The second condition further constrains the derivability relation. Both T and C are required for the derivation of E.

Finally, the following requirement is imposed:

$\langle T, C \rangle$ is an *explanation* of E iff

- $\langle T, C \rangle$ is a potential explanation of E
- C is true

The sentences constituting the explananda must be true. This is an empirical condition on the status of the explananda. $\langle T, C \rangle$ remains a *potential explanation* for E until C is verified.

From our logical perspective of chapter 2, the above conditions define a form of abduction. In potential explanation, we encounter the derivability requirement for the plain version ($T, C \vdash E$), plus one of the conditions for our 'explanatory' abductive style $C \not\vdash E$. The other condition that we had ($T \not\vdash E$) is not explicitly required above. It is implicit, however, since a significant singular sentence cannot be derived solely from quantified laws (which are usually taken to be conditional). An earlier major abductive requirement that seems absent is consistency ($T \not\vdash \neg C$). Our reading of Hempel is that this condition is certainly presupposed. Inconsistencies certainly never count as scientific explanations. Finally, the D-N account does not require minimality for explanations: it relocates such issues to choices between better or worse explanations, which fall outside the scope of the model. We have advocated the same policy for abduction in general (leaving minimal selection to our algorithms of chapter 3).

There are also differences in the opposite direction. Unlike our abductive notions of chapter 2, the D-N account crucially involves restrictions on the form of the explanantia. Also, the truth requirement is a major difference. Nevertheless, it fits well with our discussion of Peirce in chapter 1: an abducible has the status of a suggestion until it is verified.

It seems clear that the Hempelian deductive model of explanation is abductive, in that it complies with most of the logical conditions discussed in chapter 2. If we fit D-N explanation into a deductive format, the first thing to notice is that laws and initial conditions play different roles in explanation. Therefore, we need a ternary format of notation: $T \mid C \Rightarrow_{HD} E$. A consequence of this fact is that this inference is non-symmetric ($T \mid C \Rightarrow_{HD} E \not\Leftrightarrow C \mid T \Rightarrow_{HD} E$). Thus, we keep T and C separate in our further discussion. Here is the resulting notion once more:

Hempelian Deductive-Nomological Inference $\Rightarrow_{HD}$

$T \mid C \Rightarrow_{HD} E$ iff

(i) $T, C \vdash E$

(ii) T, C is consistent

(iii) $T \not\vdash E$, $C \not\vdash E$

(iv) T consists of universally quantified sentences,

C has no quantifiers or variables.

4.2.2 Structural Analysis

We now analyze this notion once more in terms of structural rules. For a general motivation of this method, see chapter 2. We merely look at a number of crucial rules discussed earlier, which tell us what kind of explanation patterns are available, and more importantly, how different explanations may be combined.

Reflexivity

Reflexivity is one form of the classical Law of Identity: every statement implies itself. This might assume two forms in our ternary format:

$$E \mid C \Rightarrow_{HD} E \qquad T \mid E \Rightarrow_{HD} E$$

However, Hempelian inference rejects both, as they would amount to 'irrelevant explanations'. Given condition (iii) above, neither the phenomenon E should count as an explanation for itself, nor should the theory contain the observed phenomenon: because no explanation would then be needed in the first place. (In addition, left reflexivity violates condition (iv), since E is not a universal but an atomic formula.) Thus, Reflexivity has no place in a structural account of explanation.

Monotonicity

Monotonic rules in scientific inference provide means for making additions to the theory or the initial conditions, while keeping scientific arguments valid. Although deductive inference by itself is monotonic, the additional conditions on $\Rightarrow_{HD}$ invalidate classical forms of monotonicity, as we have shown in detail in chapter 2. So, we have to be more careful when 'preserving explanations', adding a further condition. Unfortunately, some of the tricks from chapter 2 do not work in this case, because of our additional language requirements. Moreover, outcomes can be tricky. Consider the following monotonicity rule, which looks harmless:

HD Monotonicity on the Theory:

$$\frac{T|A \Rightarrow_{HD} E \quad T, B|D \Rightarrow_{HD} E}{T, B|A \Rightarrow_{HD} E}$$

This says that, if we have an explanation A for E from a theory, as well as another explanation for the same fact E from a strengthened theory, then the original explanation will still work in the strengthened theory. This sounds convincing, but it will fail if the strengthened theory is inconsistent with A. Indeed, we have not been able to find any convincing monotonicity rules at all!

What we learn here is a genuine limitation of the approach in chapter 2. With notions of D-N complexity, pure structural analysis may not be the best way to go. We might also just bring in the additional conditions *explicitly*, rather than encoding them in abductive sequent form. Thus, 'a theory may be strengthened in an explanation, provided that this happens consistently, and without implying the observed phenomenon without further conditions'. It is important to observe that the complexities of our analysis are no pathologies, but rather reflect the true state of affairs. They show that there is much more to the logic of Hempelian explanation than might be thought, and that this can be brought out in a precise manner. For instance, the failure of monotonicity rules means that one has to be very careful, *as a matter of logic*, in 'lifting' scientific explanations to broader settings.

Cut

The classical Cut rule allows us to chain derivations, and replace temporary assumptions by further premises implying them. Thus, it is essential to standard reasoning. Can explanations be chained? Again, there are many possible cut rules, some of which affect the theory, and some the conditions. We consider one natural version:

HD Cut:

$$\frac{T|A \Rightarrow_{HD} B \quad T|B \Rightarrow_{HD} E}{T|A \Rightarrow_{HD} E}$$

Our rain example of chapter 1 gives an illustration of this rule. Nimbostratus clouds (A) explain rain (B), and rain explains wetness (E), therefore nimbostratus clouds (A) explain wetness (E). But, is this principle generally valid? It almost looks that way, but again, there is a catch. In case A implies E by itself, the proposed conclusion does not follow. Again, we would have to add this constraint separately.

Logical Rules

A notion of inference with as many side conditions as $\Rightarrow_{HD}$ has, considerably restricts the forms of valid structural rules one can get. Indeed, this observation shows that there are clear limits to the utility of the purely structural rule analysis, which has become so popular in a broad field of contemporary logic. To get at least some interesting logical principles which govern explanation, we must bring in logical connectives. Here are some valid rules.

1. Disjunction of Theories

$$\frac{T_1 \mid C \Rightarrow E \qquad T_2 \mid C \Rightarrow E}{T_1 \vee T_2 \mid C \Rightarrow E}$$

2. Conjunction of two Explananda

$$\frac{T \mid C \Rightarrow E_1 \qquad T \mid C \Rightarrow E_2}{T \mid C \Rightarrow E_1 \wedge E_2}$$

3. Conjunction of Explanandum and Theory

$$\frac{T \mid C \Rightarrow E}{T \mid C \Rightarrow T \wedge E}$$

4. Disjunction of Explanans and Explanandum

$$\frac{T \mid C \Rightarrow E}{T \mid C \vee E \Rightarrow E}$$

5. Weakening Explanans by Theory

$$\frac{T, F \mid A \Rightarrow E}{T, F \mid F \rightarrow A \Rightarrow E}$$

The first rules shows that well-known classical inferences for disjunction and conjunction carry over to explanation. The third says that explananda can be strengthened with any consequence of the background theory. The fourth shows how explananda can be weakened 'up to the explanandum'. The fifth rule states that explananda may be weakened provided that the background theory can compensate for this. The last rule actually highlights a flaw in Hempel's models, which he himself recognized. It allows for a certain trivialization of 'minimal explanations', which might be blocked again by imposing further syntactic restrictions (see [Hem65, Sal90, p.277]).

More generally, it may be said that the D-N model has been under continued criticism through the decades after its emergence. No generally accepted formal model of deductive explanation exists. But at least we hope that our style of analysis has something fresh to offer in this ongoing debate: if only, to bring simple but essential formal problems into clearer focus.

4.2.3 The Inductive-Statistical Model

Hempel's I-S model for explaining particular events E has essentially the same form as the D-N model. The fundamental difference is the status of the laws. While in the D-N model, laws are universal generalizations, in the I-S model they are statistical regularities. This difference is reflected in the outcomes. In the I-S model, the phenomenon E is only derived 'with high probability' $[r]$ relative to the explanatory facts:

$$\begin{array}{l} L_1, \ldots, L_m \\ \underline{C_1, \ldots, C_n} \\ \hline\hline E \end{array} \; [\mathrm{r}]$$

In this schema, the double line expresses that the inference is statistical rather than deductive. This model retains all adequacy conditions of the D-N model. But it adds a further requirement on the statistical laws, known as *maximal specificity* (RMS). This requirement responds to a problem which Hempel recognized as *the ambiguity of I-S explanation*. As opposed to classical deduction, in statistical inference, it is possible to infer contradictory conclusions from consistent premises. One of our examples from chapter 1 demonstrates this.

The Ambiguity of I-S Explanation

Suppose that theory T makes the following statements. "Almost all cases of streptococcus infection clear up quickly after the administration of penicillin (L1). Almost no cases of penicillin-resistant streptococcus infection clear up quickly after the administration of penicillin (L2). Jane Jones had streptococcus infection (C1). Jane Jones received treatment with penicillin (C2). Jane Jones had a penicillin-resistant streptococcus infection (C3)." From this theory it is possible to construct two *contradictory arguments*, one explaining why Jane Jones recovered quickly (E), and the other one explaining its negation, why Jane Jones did not recover quickly ($\neg E$):

Argument 1	**Argument 2**
L_1	L_2
$\underline{C_1, C_2}$ [r]	$\underline{C_2, C_3}$ [r]
E	$\neg E$

The premises of both arguments are consistent with each other, they could all be true. However, their conclusions contradict each other, making these arguments rival ones. Hempel hoped to solve this problem by forcing all statistical laws in an argument to be *maximally specific*. That is, they should contain all relevant information with respect to the domain in question. In our example, then, premise C3 of the second argument invalidates the first argument, since the law L1 is not maximally specific with respect to all information about Jane in T. So, theory T can only explain $\neg E$ but not E.

The RMS makes the notion of I-S explanation relative to a knowledge situation, something described by Hempel as 'epistemic relativity'. This requirement helps, but it is neither a definite nor a complete solution. Therefore, it has remained controversial[1].

These problems may be understood in logical terms. Conjunction of Consequents was a valid principle for D-N explanation. It also seems a reasonable principle for explanation generally. But its implementation for I-S explanation turns out to be highly non-trivial. The RMS may be formulated semi-formally as follows:

> **Requirement of Maximal Specificity**:
>
> A universal statistical law $A \leadsto B$ is maximally specific iff for all A' such that $A' \subset A$, $A' \leadsto B$.

We should note however, that while there is consensus of what this requirement means on an intuitive level, there is no agreement as to its precise formalization

[1] One of its problems is that it is not always possible to identify a maximal specific law given two rival arguments. Examples are cases where the two laws in conflict have no relation whatsoever, as in the following example, due to Stegmúller [Ste83]: Most philosophers are not millionaires. Most mine owners are millionaires. John is both a philosopher and a mine owner. Is he a millionaire?

(cf. [Sal90] for a brief discussion on this). With this caveat, we give a version of I-S explanation in our earlier format, presupposing some underlying notion $\Rightarrow_i$ of inductive inference.

Hempelian Inductive Statistical Inference $\Rightarrow_{HI}$

$T, C \Rightarrow_{HI} E$ iff

(i) $T, C \Rightarrow_i E$

(ii) T, C is consistent

(iii) $T \not\Rightarrow_i E$, $C \not\Rightarrow_i E$

(iv) T is composed of statistical quantified formulas (which may include forms like "Most A are B"). C has no quantifiers.

(v) RMS: All laws in T are maximally specific with respect to T,C.

The above formulation complies with our earlier abductive requirements, but the RMS further complicates matters. Moreover, there is another source of vagueness. Hempel's D-N model fixes predicate logic as its language for making form distinctions, and classical consequence as its underlying engine. But in the I-S model the precise logical nature of these ingredients is left unspecified.

4.2.4 Some Interpretations of $\Rightarrow_i$

Statistical inference $\Rightarrow_i$ may be understood in a qualitative or a quantitative fashion. The former might read $\Theta \Rightarrow_i \varphi$ as: *"φ is inferred from Θ with high probability"*, while the latter might read it as *"most of the Θ models are φ models"*. These are different ways of setting up a calculus of inductive reasoning.

In addition, we have similar options as to the language of our background theories. The general statistical statements $A \rightsquigarrow B$ that may occur in the background theory Θ may be interpreted as either *"The probability of B conditioned on A is close to 1"*, or as statements of the form: *"most of the A-objects are B-objects"*. We will not pursue these options here, except for noting that the last interpretation would allow us to use the theory of *generalized quantifiers*. Many structural properties have already been investigated for probabilistic generalized quantifiers (cf. [vLa91, vBe84b] for a number of possible approaches).

Finally, the statistical approach to inference might also simplify some features of the D-N model, based on ordinary deduction. In statistical inference, the D-N notion of non-derivabilty need not be a negative statement, but rather a *positive* one of inference with (admittedly) *low probabilty*. This interpretation has some interesting consequences that will be discussed at the end of this chapter. It might also decrease complexity, since the notion of explanation becomes more uniformly 'derivational' in its formulation.

4.2.5 Structural Analysis, Revisited

Again, we briefly consider some structural properties of I-S explanation. Our discussion will be informal, since we do not fix any formal explanation of the key notions discussed before.

As for Reflexivity of $\Rightarrow_i$, this principle fails for much the same reasons as for D-N explanation. Next, consider Monotonicity. This time, new problems arise for strengthening theories in explanations, due to the RMS. A law L might be maximally specific with respect to $T|A$, but not necessarily so with respect to $T, B|A$ or to $T|A, B$. Worse still, adding premises to a statistical theory may reverse previous conclusions! In the above penicillin example, the theory without C3 explains perfectly why Jane Jones recovered quickly. But adding C3 reverses the inference, explaining instead why she did not recover quickly. If we then add that she actually took some homeopathic medicine with high chances of recovery (cf. chapter 1), the inference reverses again, and we will be able to explain once more why she recovered quickly.

These examples show once again that inductive statistical explanations are epistemically relative. There is no guarantee of preserving consistency when changing premises, therefore making monotonicity rules hopeless. And stating that the additions must be 'maximally specific' seems to beg the question. Nevertheless, we can salvage some monotonicity, provided that we are willing to *combine* deductive and inductive explanations. (There is no logical reason for sticking to pure formulations.) Here is a principle that we find plausible, modulo further specification of the precise inductive inference used (recall that $\Rightarrow_{HD}$ stands for deductive–nomological inference and $\Rightarrow_{HI}$ for inductive–statistical):

Monotonicity on the theory:

$$\frac{T|A \Rightarrow_{HI} C \quad T, B|D \Rightarrow_{HD} C}{T, B|A \Rightarrow_{HI} C}$$

This rule states that statistical arguments are monotonic in their background theory, at least when what is added explains the relevant conclusion deductively with some other initial condition. This would indeed be a valid formulation, provided that we can take care of consistency for the enlarged theory $T, B \mid A$. In particular, all maximal specific laws for $T \mid A$ remain specific for $T, B \mid A$. For, by inferring C in a deductive and consistent way, there is no place to add something that would reverse the inference, or alter the maximal specificity of the rules.

Here is a simple illustration of this rule. If on the one hand, Jane has high chances of recovering quickly from her infection by taking penicillin, and on the other she would recover by taking some medicine providing a sure cure (B), she still has high chances of recovering quickly when the assertion about this cure is added to the theory.

Not surprisingly, there is no obvious Cut rule in this setting either. Here it is not the RMS causing the problem, as the theory does not change, but rather the well-known fact that statistical implications are not transitive. Again, we have a proposal for a rule combining statistical and deductive explanation, which might form a substitute: The following is our proposed formulation:

Deductive cut on the explanans:

$$\frac{T|A \Rightarrow_{HI} B \quad T|B \Rightarrow_{HD} C}{T|A \Rightarrow_{HI} C}$$

Again, here is a simple illustration of this rule. If the administration of penicillin does explain the recovery of Jane with high probability, and this in turn explains deductively her good mood, penicillin explains with high probability Jane's good mood. (This reflects the well-known observation that "most A are B, all B are C" implies "most A are C". Note that the converse chaining is invalid.)

Patrick Suppes has asked whether one can formulate a more general representation theorem, of the kind we gave for consistent abduction (cf. chapter 2), which would leave room for statistical interpretations. This might account for the fluid boundary between deduction and induction in common sense reasoning. Exploring this question however, has no unique and easy route. As we have seen, it is very hard to formulate a sound monotonic rule for I-S explanation. But this failure is partly caused by the fact that this type of inference requires too many conditions (the same problem arose in the H-D model). So, we could explore a calculus for a simpler notion of probabilistic consequence, or rather work with a combination of deductive and statistical inferences. Still, we would have to formulate precisely the notion of probabilistic consequence, and we have suggested there are several (qualitative and quantitative) options for doing it. Thus, characterizing an abductive logical system that could accommodate statistical reasoning is a question that deserves careful analysis, beyond the confines of this thesis.

4.2.6 Further Relevant Connections

In this section we briefly present Salmon's notion of statistical relevance, as a sample of more sophisticated post-Hempelian views of explanation. Moreover, we briefly discuss a possible computational implementation of Hempel's models more along the lines of our chapter 3, and finally we present Thagard's notion of explanation within his account of 'computational philosophy of science'.

Salmon's Statistical Relevance

Despite the many conditions imposed, the D-N model still allows explanations irrelevant to the explanandum. The problem of characterizing when an explanans is *relevant* to an explanandum is a deep one, beyond formal logic. It is a key issue in the general philosophy of science.

One noteworthy approach regards relevance as *causality*. W. Salmon first analyzed explanatory relevance in terms of statistical relevance [Sal71]. For him, the issue in inductive statistical explanation is not how probable the explanans ($T|A$) renders the explanandum C, but rather whether the facts cited in the explanans make a difference to the probability of the explanandum. Thus, it is not high probability but statistical relevance that makes a set of premises statistically entail a conclusion.

Now recall that we said that statistical non-derivability ($\Theta \not\Rightarrow_i \varphi$) might be refined to state that "φ follows with low probability from Θ". With this reinterpretation, one can indeed measure if an added premise changes the probability of a conclusion, and thereby count as a relevant explanation. This is not all there is to causality. Salmon himself found problems with his initial proposal, and later developed a causal theory of explanation [Sal84], which was refined in [Sal94]. Here, explanandum and explananda are related through a *causal nexus* of causal processes and causal interactions. Even this last version is still controversial (cf. [Hit95]).

A Computational Account of Hempel's models?

Hempel treats scientific explanation as a given product, without dealing with the processes that produce such explanations. But, just as we did in chapter 3, it seems natural to supplement this inferential analysis with a computational search procedure for producing scientific explanations. Can we modify our earlier algorithms to do such a job? For the D-N model, indeed, easy modifications to our algorithm would suffice. (But for the full treatment of universal laws, we would need a predicate-logical version of the tableau algorithm.) For the inductive statistical case, however, semantic tableaux seem inappropriate. We would need to find a way of representing statistical information in tableaux, and then characterize inductive inference inside this framework. Some more promising formalisms for computing inductive explanations are labeled deductive systems with 'weights' by Dov Gabbay [Gab94a], and recent systems of dependence-driven qualitative probabilistic reasoning by W. Meyer Viol (cf. [Mey95]) and van Lambalgen & Alechina ([AL96]).

Thagard's Computational Explanation

An alternative route toward a procedural account of scientific explanation is taken by Thagard in [Tha88]. Explanation cannot be captured by a Hempelian syntactic structure. Instead, it is analyzed as a *process of providing understanding*, achieved through a mechanism of 'locating' and 'matching'. The model of explanation is the program **PI** ('processes of induction') which computes explanations for given phenomena by procedures such as abduction, analogy, concept formation, and generalization and then accounts for a 'best explanation' comparing those which the program is able to construct.

This approach concerns abduction in cognitive science (cf. chapter 1), in which explanation is regarded as a problem–solving activity modeled by computer programs. (The style of programming here is quite different from ours in chapter 3, involving complex modules and record structures.) This view gives yet another point of contact between contemporary philosophy of science and artificial intelligence.

4.2.7 Discussion

Our analysis has tested the logical tools developed in chapter 2 on Hempel's models of scientific explanation. The deductive-nomological model is indeed a form of abductive inference. Nevertheless, structural rules provide limited information, especially once we move to inductive-statistical explanation. In discussing this situation, we found that the proof theory of combinations of D-N and I-S explanation may actually be better-behaved than either by itself. Even in the absence of spectacular positive insights, we can still observe that the abductive inferential view of explanation does bring it in line with modern non-monotonic logics.

Our negative structural findings also raise interesting issues by themselves. From a logical point of view, having an inferential notion like $\Rightarrow_{HD}$ without reflexivity and (even modified) monotonicity, challenges a claim made in [Gab85]. Gabbay argues that reflexivity, 'restricted monononicity' and cut are the three minimal conditions which any consequence relation should satisfy to be a *bona fide* non-monotonic logic. (Later in [Gab94b] however, Gabbay admits this is not such a strong approach after all, as *'Other systems, such as relevance logic, do not satisfy even reflexivity'.*) We do not consider this failure an argument against the inferential view of explanation. As we have suggested in chapter 2, there are no universal logical structural rules that fit every consequence relation. What counts rather is that such relations 'fit a logical format'.

Non-monotonic logics have been mainly developed in AI. As a further point of interest, we mention that the above ambiguity of statistical explanation also occurs in default reasoning. The famous example of 'Quakers and Pacifists' is just another version of the one by Stegmüller cited earlier in this chapter. More specifically, one of its proposed solutions by Reiter [Rei80] is in fact a variation of the RMS[2]. These cases were central in Stegmüller criticisms of the positivist view of scientific explanation. He concluded that there is no satisfactory analysis using logic. But these very same examples are a source of inspiration to AI researchers developing non-standard logics.

There is a lot more to these connections than what we can cover in this dissertation. With a grain of salt, contemporary logic-based AI research may be viewed as logical positivism 'pursued by other means'. One reason why this works better nowadays than in the past, is that Hempel and his contemporaries thought that classical logic was *the logic* to model and solve their problems. By present lights, it may have been their logical apparatus more than their research program which hampered them. Even so, they also grappled with significant problems, that are as daunting to modern logical systems as to classical ones, including a variety of pragmatic factors. In all, there seem to be ample reasons for philosophers of science, AI researchers, and logicians for making common cause.

Analyzing scientific reasoning in this broader perspective also sheds light on the limitations of this dissertation. We have not really accounted for the distinction between laws and individual facts, we have no good account of 'relevance', and we have not really fathomed the depths of probability. Moreover, much of scientific explanation involves *conceptual change*, where the theory is modified with new notions in the process of accounting for new phenomena. So far, neither our logical framework

[2]In [Tan92] the author shows this fact in detail, relating current default theories to Hempel's I-S model.

of chapter 2, nor our algorithmic one of chapter 3 has anything to offer in this more ambitious realm.

4.3 Abduction as Epistemic Change

Notions related to explanation have also emerged in theories of belief change in AI. One does not just want to incorporate new beliefs, but often also, to justify them. Indeed, the work of Peter Gärdenfors, a key figure in this tradition (cf. [Gär88]) contains many explicit sources in the earlier philosophy of science. Our discussion of epistemic change will be in the same spirit, taking a number of cues from his analysis. We will concentrate on *belief revision*, where changes occur only in the theory. The situation or world to be modeled is supposed to be static, only new information is coming in. The type of epistemic change which accounts for a changing world is called *update*. We leave its connection to abduction for future research, and only briefly discuss it at the end of this chapter.

4.3.1 Theories of Belief Revision in AI

In this section, we shall expand on the brief introduction given in chapter 1, highlighting aspects that distinguish different theories of belief revision. This sets the scene for approaching abduction as a similar enterprise[3].

Given a consistent theory Θ, called the belief state, and a sentence φ, the incoming belief, there are three *epistemic attitudes* for Θ with respect to φ: either φ is accepted ($\varphi \in \Theta$), φ is rejected ($\neg\varphi \in \Theta$), or φ is undetermined ($\varphi \notin \Theta, \neg\varphi \notin \Theta$). Given these attitudes, three main operations may incorporate φ into Θ, thereby effecting an epistemic change in our currently held beliefs:

- Expansion ($\Theta + \varphi$)

  An accepted or undetermined sentence φ is added to Θ.

[3]The material of this section is mainly based on [Gär92], with some modifications taken from other approaches. In particular, in our discussion, belief revision operations are not required to handle incoming beliefs together with all their logical consequences.

- Revision ($\Theta * \varphi$):

  In order to incorporate a rejected φ into Θ and maintain consistency in the resulting belief system, enough sentences in conflict with φ are deleted from Θ (in some suitable manner) and only then is φ added.

- Contraction ($\Theta - \psi$): Some sentence ψ is retracted from Θ, together with enough sentences implying it.

Of these operations, revision is the most complex one. It may indeed be defined as a composition of the other two. First contract those beliefs of Θ that are in conflict with φ, and then expand the modified theory with sentence φ. While expansion can be uniquely defined, this is not so with contraction or revision, as several formulas can be retracted to achieve the desired effect. These operations are intuitively non-deterministic. Let me illustrate this point with example 3 from chapter 1, where we observed that the lawn is not wet, which was in conflict with our theory.

Example 3:

Θ: $r \rightarrow w$, r.

φ: $\neg w$.

In order to incorporate $\neg w$ into Θ and maintain consistency, the theory must be revised. But there are two possibilities for doing this: deleting either of $r \rightarrow l$ or r allows us to then expand the contracted theory with $\neg l$ consistently. The contraction operation per se cannot state in purely logical or set-theoretical terms which of these two options should be chosen. Therefore, an additional criterion must be incorporated in order to fix which formula to retract.

Here, the general intuition is that changes on the theory should be kept 'minimal', in some sense of informational economy. Various ways of dealing with the latter issue occur in the literature. We mention only that in [Gär88]. It is based on the notion of *entrenchment*, a preferential ordering which lines up the formulas in a belief state according to their importance. Thus, we can retract those formulas which are 'least entrenched' first. In practice, however, full-fledged AI systems of belief revision can be quite diverse. Here are some aspects that help to classify them. In our terminology, these are:

- Representation of Belief States
- Operations for Belief Revision
- Epistemological Stance

Regarding the first, we find there are essentially three ways in which the background knowledge Θ is represented: (i) belief sets, (ii) belief bases, or (iii) possible world models. A belief set is a set of sentences from a logical language L closed under logical consequence. In this classical approach, expanding or contracting a sentence in a theory is not just a matter of addition and deletion, as the logical consequences of the sentence in question should also be taken into account. The second approach emerged in reaction to the first. It represents the theory Θ as a *base for a belief set* B_Θ, where B_Θ is a finite subset of Θ satisfying $Cons(B_\Theta) = \Theta$. (That is, the set of logical consequences of B_Θ is the classical belief state). The intuition behind this is that some of the agent's beliefs have no independent status, but arise only as inferences from more basic beliefs. Finally, the more semantic approach (iii) moves away from syntactic structure, and represents theories as sets W_Θ of possible worlds (i.e., their models). Various equivalences between these approaches have been established in the literature (cf. [GR95]).

As for the second aspect, operations of belief revision can be given either 'constructively' or merely via 'postulates'. The former approach is more appropriate for algorithmic models of belief revision, the latter serves as a logical description of the properties that any such operations should satisfy. The two can also be combined. An algorithmic contraction procedure may be checked for correctness according to given postulates. (Say, one which states that the result of contracting Θ with φ should be included in the original state $(\Theta - \varphi \subseteq \Theta.)$)

Finally, each approach takes an 'epistemological stance' with respect to justification of the incoming beliefs. Here are two major paradigms. A 'foundationalist' approach argues one should keep track of the justification for one's beliefs, whereas a 'coherentist' perspective sees no need for this, as long as the changing theory stays consistent and keeps its overall coherence.

Therefore, each theory of epistemic change may be characterized by its representation of belief states, its description of belief revision operations, and its stand on the main properties of belief one should be looking for. These choices may be interdependent. Say, a constructive approach might favor a representation by belief bases, and hence define belief revision operations on some finite base, rather than the whole background theory. Moreover, the epistemological stance determines what constitutes *rational epistemc change*. The foundationalist accepts only those beliefs which are justified, thus having an additional challenge of computing the reasons for an incoming belief. On the other hand, the coherentist must maintain coherence, and hence make only those minimal changes which do not endanger (at least) consistency.

In particular the theory proposed in [Gär88] (known as the AGM paradigm after its original authors (Alchourrón, Gärdenfors, and Makinson), represents belief states as theories closed under logical consequence, while providing 'rationality postulates' to characterize the belief revision operations, and finally, it advocates a coherentist view. The latter is based on the empirical claim that people do not keep track of justifications for their beliefs, as some psychological experiments seem to indicate [Har65].

4.3.2 Abduction as Belief Revision

Abductive reasoning may be seen as an epistemic process for belief revision. In this context an incoming sentence φ is not necessarily an observation, but rather a belief for which an explanation is sought. Existing approaches to abduction usually do not deal with the issue of incorporating φ into the set of beliefs. Their concern is just how to give an account for φ. If the underlying theory is closed under logical consequence, however, then φ should be automatically added once we have added its explanation (which a foundationalist would then keep tagged as such).

Practical connections of abduction to theories of belief revision have often been noted. Of many references in the literature, we mention [AD94] (which uses abductive procedures to realize contractions over theories with 'immutability conditions'), and [Wil94] (which studies the relationship between explanations based on abduction and 'Spohnian reasons').

Our claim will be stronger. Abduction can function in a model of theory revision as a means of producing explanations for incoming beliefs. But also more generally, abductive reasoning as defined in this dissertation, itself provides a model for epistemic change. Let us discuss some reasons for this, recalling our architecture of chapter 1.

First, what were called the two 'triggers' for abductive reasoning correspond to the two epistemic attitudes of a formula being undetermined or rejected. We did not consider accepted beliefs, since these do not call for explanation.

- φ is a novelty $(\Theta \not\models \varphi, \Theta \not\models \neg\varphi) \Longleftarrow \varphi$ is undetermined
- φ is an anomaly $(\Theta \not\models \varphi, \Theta \models \neg\varphi) \Longleftarrow \varphi$ is rejected
- φ is an accepted belief $(\Theta \models \varphi)$.

The epistemic attitudes are presented in [Gär88] in terms of membership (e.g., a formula φ is accepted if $\varphi \in \Theta$). We defined them in terms of entailment, since our theories are not closed under logical consequence. In our account of abduction, both a novel phenomenon and an anomalous one involved a change in the original theory. The latter calls for a revision and the former for either expansion or revision. So, the basic operations for abduction are expansion and contraction. Therefore, both epistemic attitudes and changes in them are reflected in an abductive model.

However, our main concern is not the incoming belief φ itself. We rather want to compute and add its explanation α. But since φ is a logical consequence of the revised theory, it could easily be added. Here, then, are our abductive operations for epistemic change:

- Abductive Expansion

  Given a novel formula φ for Θ, a consistent explanation α for φ is computed and then added to Θ.

- Abductive Revision

  Given a novel or an anomalous formula φ for Θ, a consistent explanation α for φ is computed, which will involve modification of the background theory Θ into some suitably new Θ'. Again, intuitively, this involves both 'contraction' and 'expansion'.

In its emphasis on explanations, our abductive model for belief revision is richer than many theories of belief revision. Admittedly, though, not all cases of belief revision involve explanation, so our greater richness also reflects our restriction to a special setting.

Once we interpret our abductive model as a theory of epistemic change, the next question is: what kind of theory? Our main motivation is to find an explanation for an incoming belief. This fact places abduction in the above foundationalist line, which requires that beliefs have a justification. Often, abductive beliefs are used by a (scientific) community, where the earlier claim that individuals do not keep track of the justifications of their beliefs does not apply. On the other hand, an important feature of abductive reasoning is maintaining consistency of the theory. Otherwise, explanations would be meaningless. Therefore, abduction is committed to the coherentist approach as well. This is not a case of opportunism. Abduction rather demonstrates that the earlier philosophical stances are not incompatible. Indeed, [Haa93] argues for an intermediate stance of 'foundherentism'. Combinations of foundationalist and coherentist approaches are also found in the AI literature (cf. [Gal92]). Moreover, taking into account the computation of explanations of incoming beliefs makes an espistemic model closer to theories of theory refinement in machine learning [SL90, Gin88].

4.3.3 Semantic Tableaux Revisited: Toward An Abductive Model for Belief Revision

The combination of stances that we just described naturally calls for a procedural approach to abduction as an activity. But then, the same motivations that we gave in chapter 3 apply. Semantic tableaux provided an attractive constructive representation of theories, and abductive expansion operations that work over them. So, here is a further challenge for this framework. Can we extend our abductive tableau procedures to also deal with revision?

What we need for this purpose is an account of contraction on tableaux. Revision will then be forthcoming through combination with expansion, as has been mentioned before.

Revision in Tableaux

Our main idea is extremely straightforward. In semantic tableaux, contraction of a theory Θ, so as to give up some earlier consequences, translates into the *opening* of a closed branch of $\mathcal{T}(\Theta)$. Let us explain this in more detail for the case of revision. The latter process starts with Θ, φ for which $\mathcal{T}(\Theta \cup \varphi)$ is closed. In order to revise Θ, the first goal is to stop $\neg\varphi$ from being a consequence of Θ. This is done by opening a closed branch of $\mathcal{T}(\Theta)$ not closed by φ, thus transforming it into $\mathcal{T}(\Theta$'). This first step solves the problem of retracting inconsistencies. The next step is (much as in chapter 3) to find an explanatory formula α for φ by extending the modified Θ' as to make it entail φ. Therefore, revising a theory in the tableau format can be formulated as a combination of two equally natural moves, namely, *opening* and *closing* of branches:

> Given Θ, φ for which $\mathcal{T}(\Theta \cup \varphi)$ is closed, α is an abductive explanation if
>
> 1. There is a set of formulas $\beta_1, \ldots, \beta_l$ ($\beta_i \in \Theta$) such that
>
> $\mathcal{T}(\Theta \cup \varphi) - (\beta_1, \ldots, \beta_l)$ is open.
>
> Moreover, let $\Theta_1 = \Theta - (\beta_1, \ldots, \beta_l)$. We also require that
>
> 2. $\mathcal{T}((\Theta_1 \cup \neg\varphi) \cup \alpha)$ is closed.

How to implement this technically? To open a tableau, it may be necessary to retract several formulas $\beta_1, \ldots, \beta_l$ and not just one[4]. The second item in this formulation is precisely the earlier process of abductive extension, which has been

[4]The reason is that sets of formulas which entail φ should be removed. E.g., given $\Theta = \{\alpha \rightarrow \beta, \alpha, \beta\}$ and $\varphi = \neg\beta$, in order to make $\Theta, \neg\beta$ consistent, one needs to remove either $\{\beta, \alpha\}$ or $\{\beta, \alpha \rightarrow \beta\}$.

developed in chapter 3. Therefore, from now on we concentrate on the first point of the above process, namely, how to contract a theory in order to restore consistency.

Our discussion will be informal. Through a series of examples, we discover several key issues of implementing contraction in tableaux. We explore some complications of the framework itself, as well as several strategies for restoring consistency, and the effects of these in the production of explanations for anomalous observations.

4.3.4 Contraction in Tableaux

The general case of contraction that we shall need is this. We have a currently inconsistent theory, of which we want to retain some propositions, and from which we want to reject some others. In the above case, the observed anomalous phenomenon was to be retained, while the throwaways were not given in advance, and must be computed by some algorithm. We start by discussing the less constrained case of any inconsistent theory, seeing how it may be made consistent through contraction, using its semantic tableau as a guide.

As is well–known, a contraction operation is not uniquely defined, as there may be several options for removing formulas from a theory Θ so as to restore consistency. Suppose $\Theta = \{p \wedge q, \neg p\}$. We can remove either $p \wedge q$ or $\neg p$ – the choice of which depends, as we have noticed, on preferential criteria aiming at performing a 'minimal change' over Θ.

We start by noting that opening a branch may not suffice for restoring consistency. Consider the following example.

Example 1

Let $\Theta = \{p \wedge q, \neg p, \neg q\}$

$$
\begin{array}{c}
\mathcal{T}(\Theta) \\
| \\
p \wedge q \\
| \\
p \\
| \\
q \\
| \\
\neg p \\
| \\
\otimes
\end{array}
$$

By removing $\neg p$ the closed branch is opened. However, note that this is not sufficient to restore consistency in Θ because $\neg q$ was never incorporated to the tableau! Thus, even upon removal of $\neg p$ from Θ, we have to 'recompute' the tableau, and we will find another closure, this time, because of $\neg q$. This phenomenon reflects a certain design decision for tableaux, which seemed harmless as long as we are merely testing for standard logical consequence. When constructing a tableau, as soon as a literal $\neg l$ may close a branch (i.e., l appears somewhere higher up; or vice versa) it does so, and no formula is added thereafter. Therefore, when opening a branch we are not sure that all formulas of the theory are represented on it. Thus, considerable reconfiguration (or even total reconstruction) may be needed before we can decide that a tableau has 'really' been opened. Of course (for our purposes), we might change the format of tableaux, and compute closed branches 'beyond inconsistency', so as to make all sources of closure explicit.

'Recomputation' is a complication arising from the specific tableau framework that we use, suggesting that we need to do more work in this setting than in other approaches to abduction. Moreover, it also illustrates that 'hidden conventions' concerning tableau construction may have unexpected effects, once we use tableaux for new purposes, beyond their original motivation. Granting all this, we feel that such phenomena are of some independent interest, and we continue with further examples demonstrating what tableaux have to offer for the study of contraction and restoring consistency.

4.3.5 Global and Local: Strategies for Contraction

Consider the following variant of our preceding example:

Example 2

Let $\Theta = \{p \wedge q, \neg p\}$

$$\begin{array}{c} \mathcal{T}(\Theta) \\ | \\ p \wedge q \\ | \\ p \\ | \\ q \\ | \\ \neg p \\ | \\ \otimes \end{array}$$

In this tableau, we can open the closed branch by removing either $\neg p$ or p. However, while $\neg p$ is indeed a formula of Θ, p is not. Here, if we follow standard accounts of contraction in theories of belief revision, we should trace back the Θ-source of this subformula ($p \wedge q$ in this case) and remove it. But tableaus offer another route. Alternatively, we could explore 'removing subformulas' from a theory by merely modifying their source formulas, as a more delicate kind of minimal change. These two alternatives suggest two strategies for contracting theories, which we label *global* and *local* contraction, respectively. Notice, in this connection, that each occurrence of a formula on a branch has a *unique history* leading up to one specific corresponding subformula occurrence in some formula from the theory Θ being analyzed by the tableau.

The following illustrates what each strategy outputs as the contracted theory in our example:

- **Global Strategy**

  Branch–Opening $=$ $\{\neg p, p\}$

  (i) Contract with $\neg p$:

  $\neg p$ corresponds to $\neg p$ in Θ

  Θ'$=\Theta - \{\neg p\} = \{p \wedge q\}$

  (ii) Contract with p:

  p corresponds to $p \wedge q$ in Θ

  Θ' $= \Theta - \{p \wedge q\} = \{\neg p\}$

- **Local Strategy**

  Branch–Opening $=$ $\{\neg p, p\}$

  (i) Contract with $\neg p$

  Replace in the branch all connected occurrences of $\neg p$ (following its upward history) by the atom 'true': T.

$$\begin{array}{c} \mathcal{T}(\Theta) \\ | \\ p \wedge q \\ | \\ p \\ | \\ q \\ | \\ T \\ | \\ \bigcirc \end{array}$$

$\Theta' \quad = \quad \Theta - \{\neg p\} = \{p \wedge q, T\}$

(ii) Contract with p:

Replace in the branch all connected occurrences of p (following its history) by the atom true: T.

$$\begin{array}{c} \mathcal{T}(\Theta) \\ | \\ T \wedge q \\ | \\ T \\ | \\ q \\ | \\ \neg p \\ | \\ \bigcirc \end{array}$$

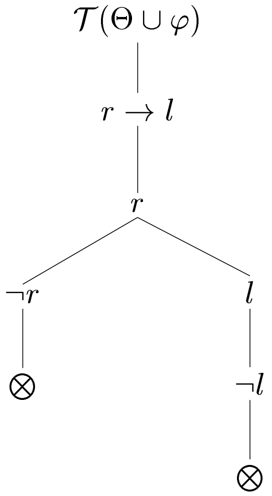
$\Theta' \quad = \quad \Theta - \{p\} = \{T \wedge q, \neg p\}$

Here, we have a case in which the two strategies differ. When contracting with p, the local strategy gives a revised theory (which is equivalent to $\{q, \neg p\}$) with less change than the global one. Indeed, if p is the source of inconsistency, why remove the whole formula $p \wedge q$ when we could modify it by $T \wedge q$? This simple example shows that 'removing subformulas' from branches, and modifying their source formulas, gives a more minimal change than removing the latter.

However, the choice is often less clear-cut. Sometimes the local strategy produces contracted theories which are logically equivalent to their globally contracted counterparts. Consider the following illustration (a variation on example 3 from chapter 1).

Example 3:

$\Theta = \{r \rightarrow l, r, \neg l\}$.



Again, we briefly note the obvious outcomes of both local and global contraction strategies.

- **Global Strategy**

  Left Branch–Opening $=$ $\{r, \neg r\}$

  (i) Contract with r:

  $\Theta' = \Theta - \{r\} = \{r \to l, \neg l\}$

  (ii) Contract with $\neg r$:

  $\Theta' = \Theta - \{r \to l\} = \{r, \neg l\}$

  Right Branch–Opening $=$ $\{\neg l, l\}$

  (i) Contract with l:

  $\Theta' = \Theta - \{r \to l\} = \{r, \neg l\}$

  (ii) Contract with $\neg l$:

  $\Theta' = \Theta - \{\neg l\} = \{r, r \to l\}$

- **Local Strategy**

  Left Branch–Opening $=$ $\{r, \neg r\}$

  (i) Contract with r:

$$\Theta\text{'} = \Theta - \{r\} = \{r \rightarrow l, \neg l\}$$

(ii) Contract with $\neg r$:

$$\Theta\text{'} = \{T \vee l, r, \neg l\}.$$

Right Branch–Opening $=$ $\{r, \neg r\}$

(i) Contract with l:

$$\Theta\text{'} = \{\neg r \vee T, r, \neg l\}$$

(ii) Contract with $\neg l$:

$$\Theta\text{'} = \Theta - \{\neg l\} = \{r, r \rightarrow l\}$$

Now, the only deviant case in the local strategy is this. Locally contracting Θ with $\neg r$ makes the new theory $\{T \vee l, r, \neg l\}$. Given that the first formula is a tautology, the output is logically equivalent to its global counterpart $\{r, \neg l\}$. Therefore, modifying versus deleting conflicting formulae makes no difference in this whole example.

A similar indifference shows up in computations with simple disjunctions, although more complex theories with disjunctions of conjunctions may again show differences between the two strategies. We refrain from spelling out these examples here, which can easily be supplied by the reader. Also, we leave the exact domain of equivalence of the two strategies as an open question. Instead, we survey a useful practical case, again in the form of an example.

4.3.6 Computing Explanations

Let us now return to abductive explanation for an anomaly. We want to keep the latter fixed in what follows (it is precisely what needs to be accommodated), modifying merely the background theory. As it happens, this constraint involves just an easy modification of our contraction procedure so far.

Example 4

$\Theta = \{p \wedge q \rightarrow r, p \wedge q\}, \qquad \varphi = \neg r$

There are five possibilities on what to retract ($\neg r$ does not count since it is the anomalous observation). In the following descriptions of 'local output', note that a removed literal $\neg l$ will lead to a substitution of F (the 'falsum') for its source l in a formula of the input theory Θ.

- Contracting with p:

  Global Strategy: Θ'=$\{p \wedge q \rightarrow r\}$

  Local Strategy: Θ'=$\{p \wedge q \rightarrow r, T \wedge q\}$

- Contracting with $\neg p$:

  Global Strategy: Θ'=$\{p \wedge q\}$

  Local Strategy: Θ'=$\{F \wedge q \rightarrow r, p \wedge q\}$

- Contracting with q:

  Global Strategy: Θ'=$\{p \wedge q \rightarrow r\}$

  Local Strategy: Θ'=$\{p \wedge q \rightarrow r, p \wedge T\}$

- Contracting with $\neg q$:

  Global Strategy: Θ'=$\{p \wedge q\}$

  Local Strategy: Θ'=$\{p \wedge F \rightarrow r, p \wedge q\}$

- Contracting with r:

  Global Strategy: Θ'=$\{p \wedge q\}$

  Local Strategy: Θ'=$\{p \wedge q \rightarrow T, p \wedge q\}$.

The only case in which the revised theories are equivalent is when contracting with r, so we have several cases in which we can compare the different explanations produced by our two strategies. To obtain the latter, we need to perform 'positive' standard abduction over the contracted theory. Let us look first at the case when the theory was contracted with p. Following the global strategy, the only explanation for $\varphi = \neg r$ with respect to the revised theory (Θ'=$\{p \wedge q \rightarrow r\}$) is the trivial solution,

$\neg r$ itself. On the other hand, following the local strategy, there is another possible explanation for $\neg r$ with respect to its revised theory (Θ'=$\{p \wedge q \rightarrow r, T \wedge q\}$), namely $q \rightarrow \neg r$. Moreover, if we contract with $\neg p$, we get the same set of possible explanations in both strategies. Thus, again, the local strategy seems to allow for more 'pointed' explanations of anomalous observations.

We do not claim that either of our strategies is definitely better than the other one. We would rather point at the fact that tableaux admit of many plausible contraction operations, which we take to be a vindication of our framework. Indeed, tableaux also suggest a slightly more ambitious approach. We outline yet another strategy to restore consistency. It addresses a point mentioned in earlier chapters, viz. that explanation often involves changing one's 'conceptual framework'.

4.3.7 Contraction by Revising the Language

Suppose we have $\Theta = \{p \wedge q\}$, and we observe or learn that $\neg p$. Following our global contraction strategy would leave the theory empty, while following the local one would yield a revised theory with $T \wedge q$ as its single formula. But there is another option, equally easy to implement in our tableau algorithms. After all, in practice, we often resolve contradictions by 'making a distinction'. Mark the proposition inside the 'anomalous formula' ($\neg p$) by some *new proposition letter* (say p'), and replace its occurrences (if any) in the theory by the latter. In this case we obtain a new consistent theory Θ' consisting of $p \wedge q, \neg p$'. And other choice points in the above examples could be marked by suitable new proposition letters as well.

We may think of the pair p, p' as two variants of the same proposition, where some distinction has been made. Here is a simple illustration of this formal manipulation.

$p \wedge q$: "Rich and Famous"

$\neg p$: "Materially poor"

$\neg p$': "Poor in spirit"

In a dialogue, the 'anomalous statement' might then be defused as follows.

– A: "X is a rich and famous person, but X is poor."

– B: Why is X poor?"

Possible answers:

– A: "Because X is poor *in spirit*"

– A: "Because being rich makes X poor *in spirit*"

– A: "Because being famous makes X poor *in spirit*"

Over the new contracted (and reformulated) theories, our abductive algorithms of chapter 3 can easily produce these three consistent explanations (as $\neg p'$, $p \rightarrow \neg p'$, $q \rightarrow \neg p'$). The idea of reinterpreting the language to resolve inconsistencies suggests that there is more to belief revision and contraction than removing or modifying given formulas. The language itself may be a source of the anomaly, and hence *it* needs revision, too. (For a more delicate study of linguistic change in resolving paradoxes and anomalies, cf. [Wei79]. For related work in argumentation theory, cf. [vBe94].) This might be considered as a simple case of 'conceptual change'.

What we have shown is that, at least some language changes are easily incorporated into tableau algorithms, and are even suggested by them. Intuitively, any inconsistent theory can be made consistent by introducing enough distinctions into its vocabulary, and 'taking apart' relevant assertions. We must leave the precise extent, and algorithmic content, of this 'folklore fact' to further research.

Another appealing consequence of to accommodating inconsistencies via language change, concerns structural rules of chapter 2. We will only mention that structural rules would acquire yet another parameter in their notation, namely the vocabulary over which the formulas are to be interpreted (eg. $p|V_1, q|V_2 \Rightarrow p \wedge q|V_1 \cup V_2$). Interestingly, this format was also used in Bolzano [Bol73]. An immediate side effect of this move are refined notions of consistency, in the spirit of those proposed by Hofstadter in [Hof79], in which consistency is relative to an 'interpretation'.

4.3.8 Outline of Contraction Algorithms

Global Strategy

Input: Θ, φ for which $\mathcal{T}(\Theta \cup \varphi)$ is closed.

Output: Θ' (Θ contracted) for which $\mathcal{T}(\Theta' \cup \varphi)$ is open.

Procedure: CONTRACT($\Theta, \neg\varphi, \Theta'$)

Construct $\mathcal{T}(\Theta \cup \varphi)$, and label its closed branches: $\Gamma_1, \ldots, \Gamma_n$.

- IF $\neg\varphi \notin \Theta$

  Choose a closed branch Γ_i (not closed by $\varphi, \neg\varphi$; if there are none, then choose any open branch).

  Calculate the literals which open it: Branch–Opening(Γ_i) $= \{\gamma_1, \gamma_2\}$. Choose one of them, say $\gamma = \gamma_1$.

  Find a corresponding formula γ' for γ in Θ higher up in the branch (γ' is either γ itself, or a formula in conjunctive or disjunctive form in which γ occurs.)

  Assign $\Theta' \quad := \quad \Theta - \gamma'$.

- ELSE ($\neg\varphi \in \Theta$)

  Assign $\Theta' \qquad = \qquad \Theta - \varphi$.

- IF $\mathcal{T}(\Theta' \cup \neg\varphi)$ is open AND all formulas from Θ are represented in the open branch, then go to END.

- ELSE

  IF $\mathcal{T}(\Theta' \cup \neg\varphi)$ is OPEN

  Add remaining formulas to the open branch(es) until there are no more formulas to add or until the tableau closes. If the resulting tableau Θ" is open, reassign Θ':=Θ" and goto END, else CONTRACT(Θ",$\neg\varphi, \Theta$"').

  ELSE CONTRACT(Θ',$\neg\varphi, \Theta$").

  % (This is the earlier-discussed 'iteration clause' for tableau recomputation.)

- END

  % (Θ' is the contracted theory with respect to $\neg\varphi$ such that $\mathcal{T}(\Theta' \cup \varphi)$ is open.)

Local Strategy

Input: Θ, φ for which $\mathcal{T}(\Theta \cup \varphi)$ is closed.

Output: Θ' (Θ contracted) for which $\mathcal{T}(\Theta' \cup \varphi)$ is open.

Procedure: CONTRACT($\Theta, \neg\varphi, \Theta'$)

Construct $\mathcal{T}(\Theta \cup \varphi)$, and label its closed branches: $\Gamma_1, \ldots, \Gamma_n$.

- Choose a closed branch Γ_i (not closed by $\varphi, \neg\varphi$; if there are none, then choose any open branch).

  Calculate the literals which open it: Branch–Opening(Γ_i) = $\{\gamma_1, \gamma_2\}$. Choose one of them, say $\gamma = \gamma_1$.

  Replace γ by T together with all its occurrences up in the branch. (γ' is either γ itself, or a formula in conjunctive or disjunctive form in which γ occurs.)

  Assign $\Theta' \quad := \quad [T/\gamma]\Theta$.

- IF $\mathcal{T}(\Theta' \cup \neg\varphi)$ is open AND all formulas from Θ are represented in the open branch, then go to END.

- ELSE

  IF $\mathcal{T}(\Theta' \cup \neg\varphi)$ is OPEN

  Add remaining formulas to the open branch(es) until there are no more formulas to add or until the tableau closes. If the resulting tableau Θ" is open, reassign $\Theta' \quad := \quad \Theta''$ and goto END, else CONTRACT($\Theta'', \neg\varphi, \Theta'''$).

  ELSE CONTRACT($\Theta', \neg\varphi, \Theta''$).

- END

  % (Θ' is the contracted (by T substitution) theory with respect to $\neg\varphi$ such that $\mathcal{T}(\Theta' \cup \varphi)$ is open.)

4.3.9 Rationality Postulates

To conclude our informal discussion of contraction in tableaux, we briefly discuss the AGM rationality postulates. (We list these postulates at the end of this chapter.) These are often taken to be the hallmark of any reasonable operation of contraction and revision – and many papers show laborious verifications to this effect. What do these postulates state in our case, and do they make sense?

To begin with, we recall that theories of epistemic change differed (amongst other things) in the way their operations were defined. These can be given either 'constructively', as we have done, or via 'postulates'. The former procedures might then be checked for correctness according to the latter. However, in our case, this is not as straightforward as it may seem. The AGM postulates take belief states to be theories closed under logical consequence. But our tableaux analyze non-deductively closed finite sets of formulas, corresponding with 'belief bases'. This will lead to changes in the postulates themselves.

Here is an example. Postulate 3 for contraction says that: "If the formula to be retracted does not occur in the belief set K, nothing is to be retracted":

K-3 If $\varphi \notin K$, then $K - \varphi = K$.

In our framework, we cannot just replace belief states by belief bases here. Of course, the intuition behind the postulate is still correct. If φ is not a consequence of Θ (that we encounter in the tableau), then it will never be used for contraction by our algorithms. Another point of divergence is that our algorithms do not put the same emphasis on contracting one specific item from the background theory as the AGM postulates. This will vitiate further discussion of even more complex postulates, such as those breaking down contractions for complex formulas into successive cases.

One more general reason for this mismatch is the following. Despite their operational terminology (and ideology), the AGM postulates describe expansions, contractions, and revisions as (in the terminology of chapter1) epistemic *products*, rather than *processes* in their own right. This gives them a 'static' flavor, which may not always be appropriate.

Therefore, we conclude that the AGM postulates as they stand do not seem to apply to contraction and revision procedures like ours. Evidently, this raises the issue of which general features *are* present in our algorithmic approach, justifying it as a legitimate notion of contraction. There is still a chance that a relatively slight 'revision of the revision postulates' will do the job. (An alternative more 'procedural' approach might be to view these issues rather in the *dynamic logic* setting of [dRi94], [vBe96a].) We must leave this issue to further investigation.

4.3.10 Discussion and Questions

The preceding was our brief sketch of a possible use of semantic tableaux for performing all operations in an abductive theory of belief revision. Even in this rudimentary state, it presents some interesting features. For a start, expansion and contraction are not reverses of each other. The latter is essentially more complex. Expanding a tableau for Θ with formula φ merely hangs the latter to the open branches of $\mathcal{T}(\Theta)$. But retracting φ from Θ often requires complete reconfiguration of the initial tableau, and the contraction procedure needs to iterate, as a cascade of formulas may have to be removed. The advantage of contraction over expansion, however, is that we need not run any separate consistency checks, as we are merely weakening a theory.

We have not come down in favor of any of the strategies presented. The local strategy tends to retain more of the original theory, thus suggesting a more minimal change than the global one. Moreover, its 'substitution approach' is nicely in line with our alternative analysis of tableau abduction in section 3.11.2, and it may lend itself to similar results. But in many practical cases, the more standard 'global abduction' works just as well. We must leave a precise comparison to future research.

Regarding different choices of formulas to be contracted, our algorithms are blatantly non-deterministic. If we want to be more focused, we would have to exploit the tableau structure itself to represent (say) some entrenchment order. The more fundamental formulas would lie closer to the root of the tree. In this way, instead of constructing all openings for each branch, we could construct only those closest to the leaves of the tree. (These correspond to the less important formulas, leaving the inner core of the theory intact.)

It would be of interest to have a proof-theoretic analysis of our contraction procedure. In a sense, the AGM 'rationality postulates' may be compared with the structural rules of chapter 2. And the more general question we have come up against is this: what are the appropriate logical constraints on a process view of contraction, and revision by abduction?

Finally, regarding other epistemic operations in AI and their connection to abduction, we have briefly mentioned *update*, the process of keeping beliefs up–to–date as the world changes. Its connection to abduction points to an interesting area of research; the changing world might be full of new surprises, or existing beliefs might have lost their explanations. Thus, appropriate operations would have to be defined to keep the theory updated with respect to these changes.

4.3.11 Conclusions

In this second part of the chapter we gave an account of abduction as a theory of belief revision. We proposed semantic tableaux as a logical representation of belief bases over which the major operations of epistemic change can be performed. The resulting theory combines the foundationalist with the coherentist stand in belief revision.

We claimed that beliefs need justification, and used our abductive machinery to construct explanations of incoming beliefs when needed. The result is richer than standard theories in AI, but it comes at the price of increased complexity. In fact, it has been claimed in [Doy92] that a realistic workable system for belief revision must not only trade deductive closed theories for belief bases, but also drop the consistency requirement. (As we saw in chapter 2, the latter is undecidable for sufficiently expressive predicate-logical languages. And it may still requires exponential time for sentences in propositional logic.) Given our analysis in chapter 2, we claim that any system which aims at producing genuine explanations for incoming beliefs must maintain consistency. What this means for workable systems of belief revision remains a moot point.

The tableau analysis confirms the intuition that revision is more complex than expansion, and that it admits of more variation. Several choices are involved, for which there seem to be various natural options, even in this constrained logical setting. What we have not explored in full is the way in which tableaux might generate entrenchment orders that we can profit from computationally. As things stand, different tableau procedures for revision may output very different explanations: abductive revision is not one unique procedure, but a family.

Even so, the preceding analysis may have shown that the standard logical tool of semantic tableaux has more uses than those for which they were originally designed.

4.4 Explanation and Belief Revision

We conclude with a short discussion and an example which relates both themes of this chapter, abduction as scientific explanation and abduction as a model for epistemic change. Theories of belief revision are mainly concerned with common sense reasoning, while theories of explanation in the philosophy of science mainly concern scientific inquiry. Nevertheless, some ideas by Gärdenfors on explanation (chapter 8 of his book [Gär88]), turn out illuminating in creating a connection. Moreover, how explanations are computed for incoming beliefs makes a difference in the type of operation required to incorporate the belief. We give an example relating to our medical diagnosis case in the first part of this chapter, to illustrate this point.

4.4.1 Explanation in Belief Revision

Gärdenfors' basic idea is that an explanation is that which makes the explanandum less surprising by raising its probability. The relationship between explananda and explanandum is relative to an epistemic state, based on a probabilistic model involving a set of possible worlds, a set of probability measures, and a belief function. Explanations are propositions which effect a special epistemic change, increasing the belief value of the explanandum. Explananda must also convey information which is 'relevant' to the beliefs in the initial state. This proposal is very similar to Salmon's views on statistical relevance, which we discussed in connection with our statistical reinterpretation of non-derivability as derivability with low probability. The main difference between the two is this. While Gärdenfors requires that the change is in raising the probability, Salmon admits just any change in probability. Gärdenfors notion of explanation is closer to our 'partial explanations' of chapter 3 (which closed some but not all of the available open tableau branches). A natural research topic will be to see if we can implement the idea of raising probability in a qualitative manner in tableaux. Gärdenfors' proposal involves degrees of explanation, suggesting a measure for explanatory power with respect to an explanandum.

Combining explanation and belief revision into one logical endeavor also has some broader attractions. This co-existence (and possible interaction) seems to reflect actual practice better than Hempel's models, which presuppose a view of scientific progress as mere accumulation. This again links up with philosophical traditions that have traditionally been viewed as a non–logical, or even anti-logical. These include historic and pragmatic accounts [Kuh70, vFr80] that focus on analyzing explanations of anomalous instances as cases of *revolutionary* scientific change. ("Scientific revolutions are taken to be those noncumulative developmental episodes in which an older paradigm is replaced in whole or in part by an incompatible new one" [Kuh70, p.148].) In our view, the Hempelian and Kuhnian schools of thought, far from being enemies, emphasize rather two sides of the same coin.

Finally, our analysis of explanation as compassing both scientific inference and general epistemic change shows that the philosophy of science and artificial intelligence share central aims and goals. Moreover, these can be pursued with tools from logic and computer science, which help to clarify the phenomena, and show their complexities.

4.4.2 Belief Revision in Explanation

As we have shown, computing explanations for incoming beliefs gives a richer model than many theories of belief revision, but this model is necessarily more complex. After all, it gives a much broader perspective on the processes of expansion and revision. We illustrate this point by an example which goes beyond our 'conservative algorithms' of chapter 3.

In standard belief revision in AI, given an undetermined belief (our case of novelty) the natural operation for modifying a theory is expansion. The reason is that the incoming belief is consistent with the theory, so the minimal change criterion dictates that it is enough to add it to the theory. Once abduction is considered however, the explanation for the fact has to be incorporated as well, and simple theory expansion might not be always appropriate. Consider our previous example of statistical reasoning in medical diagnosis (cf. chapter 1, and 4.2.4 of this chapter), concerning the quick recovery of Jane Jones, which we briefly reproduce as follows:

$\Theta : L_1, L_2, L_3, C_1$[5]

$\varphi : E$

Given theory Θ, we want to explain why Jane Jones recovered quickly (φ). Clearly, the theory neither claims with high probability that she recovered quickly ($\Theta \not\Rrightarrow \varphi$), nor that she did not ($\Theta \not\Rrightarrow \neg\varphi$). We have a case of novelty, the observed fact is consistent with the theory. Now suppose a doctor comes with the following explanation for her quick recovery: "After careful examination, I have come to the conclusion that Jane Jones recovered quickly because although she received treatment with penicillin and was resistant, her grandmother had given her Belladonna". This is a perfectly sound and consistent explanation. However, note that having the fact that 'Jane Jones was resistant to penicillin' as part of the explanation does lower the probability of explaining her quick recovery, to the point of statistically implying the contrary. Therefore, in order to make sense of the doctor's explanation, the theory needs to be revised as well, deleting the statistical rule L_2 and replacing it with something along the following lines: "Almost no cases of penicillin-resistant streptococcus infection clear up quickly after the administration of penicillin, unless they are cured by something else" (L'_2).

Thus, we have shown with this example that for the case of novelty in statistical explanation, theory expansion may not be the appropriate operation to perform (let alone the minimal one), and theory revision might be the only way to salvage both consistency and high probability.

[5] Almost all cases of streptococcus infection clear up quickly after the administration of penicillin (L1). Almost no cases of penicillin-resistant streptococcus infection clear up quickly after the administration of penicillin (L2). Almost all cases of streptococcus infection clear up quickly after the administration of Belladonna, a homeopathic medicine (L3). Jane Jones had streptococcus infection (C1). Jane Jones recovered quickly (E).

4.5 AGM Postulates for Contraction

K-1 For any sentence ϕ and any belief set K, $K - \phi$ is a belief set.

K-2 No new beliefs occur in $K - \phi$: $K - \phi \subseteq K$.

K-3 If the formula to be retracted does not occur in the belief set, nothing is to be retracted:

If $\phi \notin K$, then $K - \phi = K$.

K-4 The formula to be retracted is not a logical consequence of the beliefs retained, unless it is a tautology:

If not $\vdash \phi$, then $\phi \notin K - \phi)$.

K-5 It is possible to undo contractions (Recovery Postulate):

If $\phi \in K$, then $K \subseteq (K - \phi) + \phi$.

K-6 The result of contracting logically equivalent formulas must be the same:

If $\vdash \phi \leftrightarrow \psi$, then $K - \phi = K - \psi$.

K-7 Two separate contractions may be performed by contracting the relevant conjunctive formula:

$K - \phi \cap K - \psi \subseteq K - \phi \wedge \psi$.

K-8 If $\phi \notin K - \phi \wedge \psi$, then $K - \phi \wedge \psi \subseteq K - \psi$.

Appendix A

Algorithms for Chapter 3

In this appendix, we give a description of our tableaux algorithms for computing plain and consistent abductions. We concentrate on the over-all structure, and main procedure calls. Moreover, an actual prolog code (written in Arity Prolog) for some of these procedures, may be found at the end of this appendix.

We begin by recalling the four basic operations for constructing abductive explanations in tableau (cf. chapter 3 for their formal definitions):

BTC : The set of literals closing an open branch.

TTC : The set of literals closing all branches at once.

BPC : The set of literals which close a branch but not all the rest.

TPC : The set of literals partially closing the tableau.

Plain Abductive Explanations

- **INPUT:**

  . The theory (Θ) is given as a set of propositional formulas, separated by commas.

  . The fact to be explained (φ) is given as a literal formula.

  . Preconditions: Θ, φ are such that $\Theta \not\models \varphi$, $\Theta \not\models \neg\varphi$.

- **OUTPUT:**

Produces the set of abductive explanations $\alpha_1, \ldots \alpha_n$ such that:

(i) $\mathcal{T}((\Theta \cup \neg\varphi) \cup \alpha_i)$ is closed.

(ii) α_i complies with the vocabulary and form restrictions (cf. chapter 3, section 5).

- **MAIN PROCEDURE**

$BEGIN$

construct-tableau($[\Theta, \neg\varphi], [], [\Gamma_1, \ldots, \Gamma_k]$),

get-open-branches($[\Gamma_1, \ldots, \Gamma_k], [\Gamma_1, \ldots, \Gamma_n]$),

construct-atomic-exps($[\Gamma_1, \ldots, \Gamma_n], Atomic - exps$),

$i := 1$, empty-sol $:= false$,

% Check if all branches have a non-empty BPC, and calculate them.

$REPEAT$

IF $BPC(\Gamma_i) = \emptyset$ $THEN$ empty-sol $:= true$,

$ELSE$ $i := i + 1$

$UNTIL$ empty-sol OR $i = n + 1$

IF not(empty-sol) $THEN$

construct-conjunctive-exps($[BPC(\Gamma_1), \ldots, BPC(\Gamma_n)], Conj - exps$),

$ELSE$ $Conj - exps := \emptyset$

construct-disjunctive-exps($Atomic - exps, Conj - exps, \varphi, Disj - exps$),

write("The following are the output abductive explanations: "),

write("Atomic Explanations: "$Atomic - exps$),

write("Conjunctive Explanations: "$Conj - exps$),

write("Disjunctive Explanations: "$Disj - exps$),

END

- **SUB-PROCEDURES:**

  % Atomic Explanations

  construct-atomic-exps($[\Gamma_1, \ldots, \Gamma_n], Atomic - exps$);

  $BEGIN$

  TTC$(\Gamma_1, \ldots, \Gamma_n) := \{\gamma_1, \ldots, \gamma_m\}$,

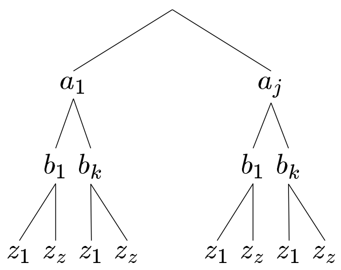
  Atomic-exps $:= \{\gamma_1, \ldots, \gamma_m\}$

  END

  % Conjunctive Explanations

  construct-conjunctive-exps($[BPC(\Gamma_1), \ldots, BPC(\Gamma_n)]$,Conj-exps);

  % Each BPC(Γ_i) contains those literals which partially close the tableau. Conjunctive explanations are constructed by taking one literal of each of these and making their conjunction. This process can be easily illustrated by the following tree structure: %



  % Each tree level represents the BPC for a branch Γ_i. (eg. BPC$(\Gamma_1) = \{a_1, \ldots, a_j\}$). Each tree branch is a conjunctive abductive solution: %

  $\beta_1 = a_1 \wedge b_1 \wedge \ldots \wedge z_1$

  $\beta_2 = a_1 \wedge b_2 \wedge \ldots \wedge z_1$

  $\beta_3 = a_1 \wedge b_3 \wedge \ldots \wedge z_1$

  $\ldots$

  $\beta_k = a_1 \wedge b_n \wedge \ldots \wedge z_z$

  $\ldots$

  $\beta_p = a_j \wedge b_k \wedge \ldots \wedge z_z$

% Some of these solutions are repeated, so further reduction is needed to get the set of abductive conjunctive explanations: %

Conj-exps=$\{\beta_1, \ldots, \beta_l\}$

% Disjunctive Explanations

construct-disjunctive-exps(Atomic-exps,Conj-exps,φ,Disj-exps);

% Disjunctive explanations are constructed by combining atomic explanations among themselves, conjunctive explanations among themselves, atomic with conjunctive, and each of atomic and conjunctive with φ.% %

(i) Atomic $\vee$ Atomic

$(\gamma_1 \vee \gamma_2), \ldots, (\gamma_1 \vee \gamma_m)$, $(\gamma_2 \vee \gamma_3), \ldots, (\gamma_2 \vee \gamma_m)$, $\ldots$, $(\gamma_m \vee \gamma_1), \ldots, (\gamma_m \vee \gamma_{m-1})$.

(ii) Conjunctive $\vee$ Conjunctive

$(\beta_1 \vee \beta_2), \ldots, (\beta_1 \vee \beta_l)$, $\ldots$, $(\beta_l \vee \beta_1), \ldots, (\beta_l \vee \beta_{l-1})$

(iii) Atomic $\vee$ Conjunctive

$(\gamma_1 \vee \beta_1), \ldots, (\gamma_1 \vee \beta_l)$, $\ldots$, $(\gamma_m \vee \beta_1), \ldots, (\gamma_m \vee \beta_l)$

(iv) Atomic $\vee$ $\varphi : (\gamma_1 \vee \varphi), \ldots, (\gamma_m \vee \varphi)$

(v) Conjunctive $\vee$ $\varphi : (\beta_1 \vee \varphi), \ldots, (\beta_l \vee \varphi)$

% The union of all these make up the set of abductive disjunctive explanations: %

Disj-exps=$\{\delta_1, \ldots, \delta_r\}$

Consistent Abductive Explanations

The input is the same as for plain abductive explanations. The output requires the additional condition of consistency:

- **OUTPUT:**

  Produces the set of abductive explanations $\alpha_1, \ldots \alpha_n$ such that (i) and (ii) from above, plus:

  (iii) $\mathcal{T}(\Theta \cup \alpha_i)$ is open.

- **MAIN PROCEDURE**

$BEGIN$

construct-tableau($[\Theta], [], [\Gamma_1, \ldots, \Gamma_k]$),

get-open-branches($[\Gamma_1, \ldots, \Gamma_k], [\Gamma_1, \ldots, \Gamma_n]$),

number-elements($[\Gamma_1, \ldots, \Gamma_n], N$),

add-formula-tableau($[\Gamma_1, \ldots, \Gamma_n], \neg\varphi, [\Gamma'_1, \ldots, \Gamma'_m]$),

number-elements($[\Gamma'_1, \ldots, \Gamma'_m], N'$),

extension-type($N, N', Type$),

IF extension-type($N, N', semi-closed$) $THEN$

construct-atomic-exps($[\Gamma'_1, \ldots, \Gamma'_n], Atomic-exps$),

construct-conjunctive-exps($[BPC(\Gamma'_1), \ldots, BPC(\Gamma'_n)], Conj-exps$),

$ELSE$

$Conj-exps = \emptyset$,

$Disj-exps = \emptyset$,

write("There are neither atomic nor conjunctive explanations "),

$ENDIF$

construct-disjunctive-exps($Atomic-exps, Conj-exps, \varphi, Temp-Disj-exps$),

eliminate-inconsistent-disjunctive-exps($Temp-Disj-exps, Disj-exps$),

Explanations :=$Atomic-exps \cup Conj-exps \cup Disj-exps$,

write("The following are the output abductive explanations: "),

write(Explanations),

END

Prolog Code

What follows is a Prolog implementation for some of the procedures above (written for Arity Prolog).

This implmentation shows that abduction is not particularly hard to use in practice (which may explain its appeal to programmers.) It may be a complex notion in general, but when well-delimited, it poses a feasible programming task.

```
% BEGIN PROLOG CODE

% CLAUSES FOR THE CONSTRUCTION OF ABDUCTIVE EXPLANATIONS

% Atomic Explanations.

% Consists of intersecting the tableau partial
% closure with the tableau totalclosure.
% atomic-explanations(Tableau,Atomic-set).

construct-atomic-exps(Tableau,Atomic-exp):-
   tableau-tot-closure(Tableau,Branches-closure,TTC),
   tableau-par-closure(Tableau,[],TPC),
   intersection([TTC,TPC],Atomic-exp).

% Conjuntive Explanations.

% This predicate first constructs the partial closures
% of the open branches and then checks that indeed every
% branch has a set of partial closures, for otherwise
% there are no conjunctive explanations. Then a tree like
% structure is formed for these partial explanations in
% which each branch is a solution. Repeated literals are
% removed from every solution, and what is left is the set
% of conjunctive explanations.

% conjunctive-explanations(Branches-closure,Atomic-exp,
                           Conjunctive-exp).
```

```
construct-conjunctive-exps(Tableau,Conj-exp):-
  open-branches(Tableau,Open-branches),
  closed-branches(Tableau,Closed-branches),
  tableau-par-closure(Open-branches,[],TPC),
  number-elements(Open-branches,N),
  number-elements(Open-branches,N'),
  N=N',
  construct-tableau(TPC,[],Solution-Paths),
  reduce-solutions(Solution-Paths,Conj-exp), !.

% In case some branch partial closure is empty,
% the result of conjunctive explanations returns
% false, meaning there are no explanations in
% conjunctive form:

construct-conjunctive-exps(Tableau,[]).


% Disjunctive Explanations.

% Here we show the construction of disjunctive
% explanations between the atomic explanations
% and the observation and between conjunctive
% explanations and observation.

cons-cond-exp(X,F,[not X v F, not F v not X]).

cons-set-cond-exp([X|R], F, [Sol1|Rest]):-
  cons-cond-exp(X,F,Sol1),
  cons-set-cond-exp(R, F, Rest).
```

```
cons-set-cond-exp([], F, Result).

disj-explanations(Atomic, Partial, Obs, [Part1, Part2]):-
  cons-set-cond-exp(Atomic, F, Part1),
  cons-set-cond-exp([X|R], F, Part2).


/* CLAUSES FOR THE CONSTRUCTION
   OF BRANCH AND TABLEAU CLOSURES. */


% Branch Total Closure (BTC): For a given
% Branch, it computes those literals which
% close it.
% branch-tot-closure(Branch,List,BTC).

branch-tot-closure([X|R],List,BTC):-
  literal(X),
  branch-tot-closure(R,[List, not X],BTC), !.

branch-tot-closure([X|R],List,BTC):-
  branch-tot-closure(R,List,BTC).

branch-tot-closure([],BTC,BTC).

% Generates the sets of branch closures
% for a given tableau.
% tableau-closure(Tableau,List,Result).

tableau-closure([X|R],List-so-far,[Clos1|Res]):-
  branch-tot-closure(X,[],Clos1),
```

```
  tableau-closure(R,Clos1,Rest).

tableau-closure([],Tableau-closure,Tableau-closure).

% Tableau Total Closure (TTC)
% Intersection of all branch total closures.

tableau-tot-closure(Tableau,Branches-closure,Tclosure):-
    tableau-closure(Tableau,[],Branches-closure),
    intersection(Branches-closure,Tclosure).


% Branch Partial Closure (BPC)

% For a given branch constructs those literals
% which close it but do not close the whole tableau.

branch-par-closure(Tableau,Branch,List,BPclosure):-
    branch-tot-closure(Branch,[],BTC),
    tableau-tot-closure(Tableau,[],TTC),
    BPclosure = BTC - TTC.


% Tableau Partial Closure (TPC)

% Constructs th union of all branch partial closures.
% tableau-par-closure(Tableau,List-so-far,TPclosure).


tableau-par-closure([X|R],List-so-far,[BPC-X|Res]):-
  branch-par-closure([X|R],X,[],BPC-X),
```

```
  tableau-par-closure(R,BPC-X,Rest).

tableau-par-closure([],TPclosure,TPclosure).


% CLAUSES FOR THE CONSTRUCTION OF
% TABLEAU AND ADDITION OF FORMULAS

% Tableau Construction
% construct-tableau(Theory,Tree-so-far,Tableau-Result)

construct-tableau([X|R],Tree-so-far,Tableau-Result):-
  add-formula-tableau(Tree-so-far,X,Tree-Result),
  construct-tableau(R,Tree-Result,Tableau-Result).

construct-tableau([],Tableau-Result,Tableau-Result).

% Adding a Formula to a Tableau:
% add-formula-tableau(Tableau,Formula,New-Tableau)

add-formula-tableau([B|R],Formula,[New-Branch|Rest]): -
    add-formula-branch(B,Formula,New-Branch),
    add-formula-tableau(R,Formula,Rest).

add-formula-tableau([],Formula,Result).


% Add formula to a Branch.
% add-formula-branch(Branch,Formula,Result)

% Case: branch is closed, nothing is to be added.
```

```
add-formula-branch(Branch,Formula,Branch):-
  closed(Branch,true), !.

% Case: A is a literal, just appended at the end
% of the branch path.
add-formula-branch(Branch,A,[Branch|A]) :- literal(A).

% Case: A^B is a conjunction, first clause is the
% condition when both are literals, the second when
% they are not.

add-formula-branch(Branch,A^B,[Branch,A,B]):-
  literal(A),literal(B).

add-formula-branch(Branch,A^B,Result):-
    add-formula-branch([Branch,A^B],A,Res),
    add-formula-branch(Res,B,Result).

% Case: AvB is a disjunction, first clause is the
% condition when both are literals, the second when
% they are not.

add-formula-branch(Branch,AvB,[[Branch,A],[Branch,B]]):-
    literal(A),literal(B).

add-formula-branch(Branch,AvB,[Branch1,Branch2]):-
    add-formula-branch([Branch,AvB],A,Branch1),
    add-formula-branch([Branch,AvB],B,Branch2).

% END OF PROLOG CODE
```

Abstract

In this dissertation I study abduction, that is, reasoning from an observation to its possible explanations, from a logical point of view. This approach naturally leads to connections with theories of explanation in the philosophy of science, and to computationally oriented theories of belief change in Artificial Intelligence.

Many different approaches to abduction can be found in the literature, as well as a bewildering variety of instances of explanatory reasoning. To delineate our subject more precisely, and create some order, a general taxonomy for abductive reasoning is proposed in chapter 1. Several forms of abduction are obtained by instantiating three parameters: the kind of reasoning involved (e.g., deductive, statistical), the kind of observation triggering the abduction (novelty, or anomaly w.r.t. some background theory), and the kind of explanations produced (facts, rules, or theories). In chapter 2, I choose a number of major variants of abduction, thus conceived, and investigate their logical properties. A convenient measure for this purpose are so-called 'structural rules' of inference. Abduction deviates from classical consequence in this respect, much like many current non-monotonic consequence relations and dynamic styles of inference. As a result we can classify forms of abduction by different structural rules. A more computational analysis of processes producing abductive inferences is then presented in chapter 3, using the framework of semantic tableaux. I show how to implement various search strategies to generate various forms of abductive explanations.

Our eventual conclusion is that abductive processes should be our primary concern, with abductive inferences their secondary 'products'. Finally, chapter 4 is a confrontation of the previous analysis with existing themes in the philosophy of science and artificial intelligence. In particular, I analyse two well-known models for scientific explanation (the deductive-nomological one, and the inductive-statistical one) as forms of abduction. This then provides them with a structural logical analysis in the style of chapter 2. Moreover, I argue that abduction can model dynamics of belief revision in artificial intelligence. For this purpose, an extended version of the semantic tableaux of chapter 3 provides a new representation of the operations of expansion, and contraction.

Bibliography

[AAA90] *Automated Abduction. Working Notes*. 1990 Spring Symposium Series of the AAA. Stanford University (March 27-29,1990).

[AL96] N. Alechina and M. van Lambalgen. 'Generalized quantification as substructural logic'. *Journal of Symbolic Logic*, 61(3):1006–1044. 1996.

[Ali93] A. Aliseda. *On presuppositions: how to accommodate them*. Manuscript. Philosophy Department. Stanford University. Spring 1993.

[Ali94] A. Aliseda. *Thinking Backwards: Abductive Variations*. Manuscript. Philosophy Department. Stanford University. Spring 1994.

[Ali95] A. Aliseda. 'Abductive Reasoning: From Perception to Invention'. Paper presented at the *Stanford–Berkeley Graduate Student Conference*. Stanford University, April 1995.

[Ali96a] A. Aliseda. 'A unified framework for abductive and inductive reasoning in philosophy and AI', in *Abductive and Inductive Reasoning Workshop Notes*. pp 1–6. European Conference on Artificial Intelligence (ECAI'96). Budapest. August, 1996.

[Ali96b] A. Aliseda. 'Toward a Logic of Abduction', in *Memorias del V Congreso Iberoamericano de Inteligencia Artifical (IBERAMIA'96)*, Puebla, México (October 28th–November 1st), 1996.

[And86] D. Anderson. 'The Evolution of Peirce's Concept of Abduction'. *Transactions of the Charles S. Peirce Society 22-2:145–164*. 1986.

[And87] D. Anderson. *Creativity and the Philosophy of C.S. Peirce*. Matinus Nijhoff. Philosophy Library Volume 27. Martinus Nijhoff Publishers. 1987.

[Ant89] C. Antaki. 'Lay explanations of behaviour: how people represent and communicate their ordinary theories', in C. Ellis (ed), *Expert Knowledge and Explanation: The Knowledge–Language Interface*, pp. 201–211. Ellis Horwood Limited, England. 1989.

[AD94] C. Aravindan and P.M. Dung. 'Belief dynamics, abduction and databases'. In C. MacNish, D. Pearce, and L.M. Pereira (eds). *Logics in Artificial Intelligence. European Workshop JELIA'94*, pp. 66–85. Lecture Notes in Artificial Intelligence 838. Springer–Verlag. 1994.

[Ayi74] M. Ayim. 'Retroduction: The Rational Instinct'. *Transactions of the Charles S. Peirce Society 10-1:34–43*. 1974.

[vBe84a] J. van Benthem. 'Lessons from Bolzano'. Center for the Study of Language and Information. Technical Report CSLI-84-6. Stanford University. 1984. Later published as 'The variety of consequence, according to Bolzano'. *Studia Logica* 44, 389–403.

[vBe84b] J. van Benthem. 'Foundations of Conditional Logic'. *Journal of Philosophical Logic'* 13, 303–349. 1984.

[vBe90] J. van Benthem. 'General Dynamics'. ILLC Prepublication Series LP-90-11. Institute for Logic, Language, and Information. University of Amsterdam. 1990.

[vBe91] J. van Benthem. *Language in Action. Categories, Lambdas and Dynamic Logic*. Amsterdam: North Holland. 1991.

[vBe92] J. van Benthem. Logic as Programming. *Fundamenta Informaticae*, 17(4):285–318, 1992.

[vBe93] J. van Benthem. 'Logic and the Flow of Information'. *ILLC Prepublication Series* LP-91-10. Institute for Logic, Language, and Information. University of Amsterdam. 1993.

[vBe94] J. van Benthem. 'Logic and Argumentation'. ILLC Research Report and Technical Notes Series X-94-05. Institute for Logic, Language, and Information. University of Amsterdam. 1994.

[vBe96a] J. van Benthem. *Exploring Logical Dynamics*. CSLI Publications, Stanford University. 1996.

[vBe96b] J. van Benthem. Inference, Methodology, and Semantics. In P.I. Bystrov and V.N. Sadovsky (eds)., *Philosophical Logic and Logical Philosophy*, 63–82. Kluwer Academic Publishers. 1996.

[vBC94] J. van Benthem, G. Cepparello. 'Tarskian Variations: Dynamic Parameters in Classical Semantics'. Technical Report CS–R9419. Centre for Mathematics and Computer Science (CWI), Amsterdam. 1994.

[BE93] J. Barwise and J. Etchemendy. *Tarski's World*. Center for the Study of Language and Information (CSLI), Stanford, CA. 1993.

[Bet59] E.W. Beth. *The Foundations of Mathematics*. Amsterdam, North-Holland Pub. Co. 1959.

[Bet69] E. W. Beth. 'Semantic Entailment and Formal Derivability'. In J. Hintikka (ed), *The Philosophy of Mathematics*, p. 9–41. Oxford University Press. 1969.

[BR97] P. Blackburn and Maarten de Rijke. 'Why Combine Logics?. *Studia Logica*. To appear.

[Bol73] B. Bolzano. *Wissenschaftslehre*, Seidel Buchhandlung, Sulzbach. Translated as *Theory of Science* by B. Torrel, edited by J. Berg. D. Reidel Publishing Company. The Netherlands, Dordrecht, 1973.

[Car55] R. Carnap. *Statistical and Inductive Probability.* Galois Institute of Mathematics and Art Brooklyn, N.Y., 1955.

[Cho72] N. Chomsky. *Language and Mind. Enlarged Edition.* New York, Harcourt Brace Jovanovich. 1972.

[Cor75] J. Corcoran. 'Meanings of Implication', *Diaglos* 9, 59–76. 1975.

[Doy92] J. Doyle. 'Reason maintenance and belief revision: foundations versus coherence theories', in: P.Gärdenfors (ed), *Belief Revision*, pp. 29–51. Cambridge Tracts in Theoretical Computer Science, Cambridge University Press, 1992.

[DH93] K.Dosen and P. Schroeder-Heister (eds). *Substructural Logics.* Oxford Science Publications. Clarendon Press, Oxford. 1993.

[DP91b] D. Dubois and H. Prade. 'Possibilistic logic, preferential models, nonmonotonicity and related issues', in *Proceedings of the 12th Inter. Joint Conf. on Artificial Intelligence (IJCAI–91)*, pp. 419–424 Sidney, Australia. 1991.

[ECAI96] *Abductive and Inductive Reasoning Workshop Notes.* European Conference on Artificial Intelligence (ECAI'96). Budapest. August, 1996.

[EG95] T. Eiter and G. Gottlob. 'The Complexity of Logic-Based Abduction', *Journal of the Association for Computing Machinery*, vol 42 (1): 3–42. 1995.

[Fan70] K.T.Fann. *Peirce's Theory of Abduction.* The Hague: Martinus Nijhoff. 1970.

[Fit90] M. Fitting. *First Order Logic and Automated Theorem Proving.* Graduate Texts in Computer Science. Springer–Verlag. 1990.

[Fla95] P.A. Flach. *Conjectures: an inquiry concerning the logic of induction.* PhD Thesis, Tilburg University. 1995.

[Fla96a] P. Flach. 'Rationality Postulates for Induction', in Y. Shoham (ed), *Proceedings of the Sixth Conference on Theoretical Aspects of Rationality and Knowledge (TARK 1996)*, pp.267–281. De Zeeuwse Stromen, The Netherlands, March, 1996.

[Fla96b] P. Flach. 'Abduction and induction: syllogistic and inferential perspectives', in *Abductive and Inductive Reasoning Workshop Notes*. pp 31–35. European Conference on Artificial Intelligence (ECAI'96). Budapest. August, 1996.

[Fra58] H. Frankfurt. 'Peirce's Notion of Abduction'. *The Journal of Philosophy*, 55 (p.594). 1958.

[vFr80] B. van Frassen. *The Scientific Image*. Oxford : Clarendon Press. 1980.

[Gab85] D.M. Gabbay. 'Theoretical foundations for non-monotonic reasoning in expert systems', in K. Apt (ed), *Logics and Models of Concurrent Systems*, pp. 439–457. Springer–Verlag. Berlin, 1985.

[Gab94a] D.M Gabbay. *Labelled Deductive Systems. Part I*, Oxford University Press, 1994.

[Gab94b] D. M Gabbay. 'Classical vs Non-classical Logics (The Universality of Classical Logic)', in D.M. Gabbay, C.J. Hogger and J.A. Robinson (eds), *Handbook of Logic in Artificial Intelligence and Logic Programming. Volume 2 Deduction Methodologies*, pp.359–500. Clarendon Press, Oxford Science Publications. 1994.

[GJK94] D.M. Gabbay, R. Kempson. *Labeled Abduction and Relevance Reasoning.* Manuscript. Department of Computing, Imperial College, and Department of Linguistics, School of Oriental and African Studies. London. 1994.

[Gal92] J.L. Galliers. 'Autonomous belief revision and communication', in: P.Gärdenfors (ed), *Belief Revision*, pp. 220–246. Cambridge Tracts in Theoretical Computer Science, Cambridge University Press, 1992.

[Gär88] P. Gärdenfors. *Knowledge in Flux: Modeling the Dynamics of Epistemic States.* MIT Press. 1988.

[Gär92] P. Gärdenfors. 'Belief revision: An introduction', in: P.Gärdenfors (ed), *Belief Revision*, pp. 1–28. Cambridge Tracts in Theoretical Computer Science, Cambridge University Press, 1992.

[GR95] P. Gärdenfors and H. Rott. 'Belief revision' in in D.M. Gabbay, C.J. Hogger and J.A. Robinson (eds), *Handbook of Logic in Artificial Intelligence and Logic Programming. Volume 4*, Clarendon Press, Oxford Science Publications. 1995.

[Ger95] P. Gervás. *Logical considerations in the interpretation of presuppositional sentences.* PhD Thesis, Department of Computing, Imperial College London. 1995.

[Gie91] R. Giere (ed). *Cognitive Models of Science.* Minnesota Studies in the Philosophy of Science, vol 15. Minneapolis: University of Minnesota Press. 1991.

[Gil96] D. Gillies. *Artificial Intelligence and Scientific Method.* Oxford University Press. 1996.

[Gin88] A. Ginsberg. 'Theory Revision via prior operationalization'. In *Proceedings of the Seventh Conference of the AAAI.* 1988.

[Gro95] W. Groeneveld. *Logical Investigations into Dynamic Semantics.* PhD Dissertation, Institute for Logic, Language and Information, University of Amsterdam, 1995. (ILLC Dissertation Series 1995–18).

[Haa93] S. Haack. *Evidence and Inquiry. Towards Reconstruction in Epistemology.* Blackwell, Oxford UK and Cambridge, Mass. 1993.

[Han61] N.R. Hanson. *Patterns of Scientific Discovery.* Cambridge at The University Press. 1961.

[Har65] G. Harman. 'The Inference to the Best Explanation'. *Philosophical Review.* 74 88-95 (1965).

[Har86] G. Harman. *Change in View: Principles of Reasoning.* Cambridge, Mass. MIT Press, 1986.

[Hei83] I. Heim 'On the Projection Problem for Presuppositions', in *Proceedings of the Second West Coast Conference on Formal Linguistics, WCCFL*, 2:pp.114–26. 1983.

[Hem43] C. Hempel. 'A purely syntactical definition of confirmation', *Journal of Symbolic Logic* 6 (4):122–143. 1943.

[Hem45] C. Hempel. 'Studies in the logic of confirmation', *Mind* 54 (213): 1–26 (Part I); 54 (214): 97–121 (Part II). 1945.

[Hem65] C. Hempel. 'Aspects of Scientific Explanation', in C. Hempel *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science.* The Free Press, New York. 1965.

[HO48] C. Hempel and P. Oppenheim. 'Studies in the logic of explanation', *Philosophy of Science* 15:135–175. 1948.

[Hin55] J. Hintikka. 'Two Papers on Symbolic Logic: Form and Quantification Theory and Reductions in the Theory of Types', *Acta Philosophica Fennica*, Fasc VIII, 1955.

[Hit95] C.R. Hitchcock. 'Discussion: Salmon on Explanatory Relevance', *Philosophy of Science*, pp. 304–320, 62 (1995).

[HSAM90] J.R. Hobbs, M. Stickel, D. Appelt, and P. Martin. 'Interpretation as Abduction'. SRI International, Technical Note 499, Artificial Intelligence Center, Computing and Engineering Sciences Division, Menlo Park, Ca. 1990.

[Hof79] D.R. Hofstadter. *Gödel, Escher, Bach : an eternal golden brain.* 1st Vintage Books ed. New York : Vintage Books. 1979.

[HHN86] J. Holland, K. Holyoak, R. Nisbett, R., and P. Thagard, *Induction: Processes of inference, learning, and discovery.* Cambridge, MA:MIT Press/Bradford Books. 1986.

[Hoo92] C.Hookway. *Peirce.* Routledge, London 1992.

[Jos94] J.R. Josephson. *Abductive Inference.* Cambridge University Press, 1994.

[KM94] A. Kakas, P. Mancarella. 'Knowledge assimilation and abduction', in *Proceedings of the European Conference on Artificial Intelligence, ECAI'90.* International Workshop on Truth Maintenance, Stockholm, Springer–Verlag Lecture Notes in Computer Science (1990).

[KKT95] A.C. Kakas, R.A.Kowalski, F. Toni. 'Abductive Logic Programming', *Journal of Logic and Computation* 2(6) (1993) 719-770.

[Kal95] M. Kalsbeek. *Meta Logics for Logic Programming.* PhD Dissertation, Institute for Logic, Language and Information, University of Amsterdam, 1995. (ILLC Dissertation Series 1995–13).

[Kan93] M. Kanazawa. 'Completeness and Decidability of the Mixed Style of Inference with Composition'. Center for the Study of Language and Information. Technical Report CSLI-93-181. Stanford University. 1993.

[Kap90] T. Kapitan. 'In What Way is Abductive Inference Creative?'. *Transactions of the Charles S. Peirce Society 26/4:499–512.* 1990.

[Kon90] K. Konolige. 'A General Theory of Abduction'. In: *Automated Abduction, Working Notes*, pp. 62–66. Spring Symposium Series of the AAA. Stanford University. 1990.

[Kon96] K. Konolige. 'Abductive Theories in Artificial Intelligence', in G. Brewka (ed), *Principles of Knowledge Representation.* CSLI Publications, Stanford University. 1996.

[Kow79] R.A. Kowalski. *Logic for problem solving.* Elsevier, New York, 1979.

[Kow91] R.A. Kowalski. 'A Metalogic Programming Approach to multiagent knowledge and belief', in V. Lifschitz (ed), *Artificial Intelligence and Mathematical Theory of Computation*, pp. 231–246. Academic Press, 1991.

[KLM90] S. Kraus, D. Lehmann, M. Magidor. 'Nonmonotonic Reasoning, Preferential Models and Cumulative Logics'. *Artificial Intelligence*, 44 (1990) 167–207.

[Kru95] G. Kruijff. *The Unbearable Demise of Surprise. Reflections on Abduction in Artificial Intelligence and Peirce's Philosophy.* Master Thesis. University of Twente. Enschede, The Netherlands. August, 1995.

[Kuh70] T. Kuhn. *The Structure of Scientific Revolutions* 2nd ed. Chicago: University of Chicago Press, 1970.

[Kui87] Th.A.F. Kuipers (editor). *What is Closer-to-the-Truth?*, Rodopi, Amsterdam. 1987.

[Kur95] N. Kurtonina. *Frames and Labels. A Modal Analysis of Categorial inference.* PhD Dissertation, Institute for Logic, Language and Information. University of Amsterdam, 1995. (ILLC Dissertation Series 1995–8).

[Joh83] P.N. Johnson Laird. *Mental Models.* Cambridge, Mass. Harvard University Press, 1983.

[Lak76] I. Lakatos. *Proofs and Refutations. The Logic of Mathematical Discovery.* Cambridge Universiity Press. 1976.

[vLa91] M. van Lambalgen. 'Natural Deduction for Generalized Quantifiers' in Jaap van der Does and Jan van Eijck (eds), *Quantifiers, Logic, and Language*, vol 54 of Lecture Notes, pp 225–236. CSLI Publications, Stanford, California, 1996. Available as preprint since 1991.

[Lap04] P.S. Laplace. 'Memoires'. In *Oeuvres complétes de Laplace*, vol. 13–14. Publiéees sous les auspices de l'Académie des sciences, par MM. les secrétaires perpétuels. Paris, Gauthier-Villars, 1878-1912.

[LSB87] P. Langley and H. Simon, G. Bradshaw, and J. Zytkow. *Scientific Discovery*. Cambridge, MA:MIT Press/Bradford Books. 1987.

[LLo87] J.W. Lloyd. *Foundations of Logic Programming*. Springer–Verlag, Berlin, 2nd. edition. 1987.

[Mak93] D. Makinson. 'General Patterns in Nonmonotonic Reasoning', in D.M. Gabbay, C.J. Hogger and J.A. Robinson (eds), *Handbook of Logic in Artificial Intelligence and Logic Programming. Volume 3, Nonmonoptonic reasoning and uncertain reasoning*, chapter 2.2. Clarendon Press, Oxford Science Publications. 1994.

[MP93] M.C. Mayer and F. Pirri. 'First order abduction via tableau and sequent calculi', in *Bulletin of the IGPL*, vol. 1, pp.99–117. 1993.

[Mey95] W.P. Meyer Viol *Instantial Logic. An Investigation into Reasoning with Instances*. Ph.D Dissertation, Institute for Logic, Language and Information, University of Amsterdam, 1995. (ILLC Dissertation Series 1995–11).

[McC80] J. McCarthy. 'Circumscription: A form of non-monotonic reasoning', *Artificial Intelligence* 13 (1980) pp. 27–39.

[Men64] E. Mendelson. *Introduction to Mathematical Logic*. New York, Van Nostrand. 1964.

[Mic94] R. Michalski. 'Inferential Theory of Learning: Developing Foundations for Multistrategy Learning' in *Machine Learning: A Multistrategy Approach*, Morgan Kaufman Publishers, 1994.

[Mill 58] J.S. Mill. *A System of Logic.* (New York, Harper & brothers, 1858). Reprinted in *The Collected Works of John Stuart Mill*, J.M Robson (ed), Routledge and Kegan Paul, London.

[Min90] G.E. Mints. 'Several formal systems of the logic programming', *Computers and Artificial Intelligence*, 9:19–41. 1990.

[Nag79] E. Nagel. *The Structure of Science. Problems in the Logic of Scientific Explanation*, 2nd edition. Hackett Publishing Company, Indianapolis, Cambridge. 1979.

[Nie96] I. Niemelä. 'Implementing Circumpscription using a Tableau Method', in W. Wahlster (ed) *Proceedings of the 12th European Conference on Artificial Intelligence, ECAI'96*, pp 80–84. 1996.

[Pau93] G. Paul. 'Approaches to Abductive Reasoning: An Overview'. *Artificial Intelligence Review*, 7:109–152, 1993.

[CP] C.S. Peirce. *Collected Papers of Charles Sanders Peirce.* Volumes 1–6 edited by C. Hartshorne, P. Weiss. Cambridge, Harvard University Press. 1931–1935; and volumes 7–8 edited by A.W. Burks. Cambridge, Harvard University Press. 1958.

[PG87] D. Poole., R.G. Goebel., and Aleliunas. 'Theorist: a logical reasoning system for default and diagnosis', in Cercone and McCalla (eds), *The Knowledge Fronteer: Essays in the Representation of Knowledge*, pp. 331-352. Springer Verlag Lecture Notes in Computer Science 331-352. 1987.

[PR90] Y. Peng and J.A. Reggia. *Abductive Inference Models for Diagnostic Problem-Solving.* Springer Verlag. Springer Series in Symbolic Computation – Artificial Intelligence. 1990.

[PB83] J. Perry and J. Barwise. *Situations and Attitudes.* Cambridge, Mass: MIT Press, 1983.

[Pol45] G. Polya. *How to Solve it. A New Aspect of Mathematical Method.* Princeton University Press. 1945.

[Pol54] G. Polya. *Induction and Analogy in Mathematics.* Vol I. Princeton University Press. 1954.

[Pol62] G. Polya. *Mathematical Discovery. On Understanding, learning, and teaching problem solving.* Vol I. John Wiley & Sons, Inc. New York and London. 1962.

[Pop63] K. Popper. *Conjectures and Refutations. The Growth of Scientific Knowledge.* 5th ed. London and New York. Routledge. 1963.

[Pop58] K. Popper. *The Logic of Scientific Discovery.* London: Hutchinson. 1958.

[Pop73] H.E. Pople. On the Mechanization of Abductive Logic. In: *Proceedings of the Third International Joint Conference on Artificial Intelligence*, IJCAI-73, San Mateo: Morgan Kauffmann, Stanford, CA. pp. 147–152. 1973.

[Qui61] W.V. Quine. 'On What there is', in *From a Logical Point of View*, 2nd ed.rev. Cambridge and Harvard University Press. 1961

[Rei38] H. Reichenbach. *Experience and Prediction.* Chicago: University of Chicago Press. 1938.

[Rei70] F.E. Reilly. *Charles Peirce's Theory of Scientific Method.* New York: Fordham University Press, 1970.

[Rei80] R. Reiter. 'A Logic for default reasoning', *Artificial Intelligence* 13. 1980.

[Rei87] R. Reiter. 'A theory of diagnosis from first principles', *Artificial Intelligence* 32. 1987.

[Res78] N. Rescher. *Peirce's Philosophy of Science. Critical Studies in His Theory of Induction and Scientific Method.* University of Notre Dame. 1978.

[Rib96] P. Ribenboim. *The new book of prime number records.* 3rd edition. New York: Springer. 1996.

[dRi94] M. de Rijke. 'Meeting Some Neighbours', in J. van Eijck and A. Visser (eds) *Logic and Information Flow.* MIT Press, 1994.

[Rot88] R.J. Roth. 'Anderson on Peirce's Concept of Abduction'. *Transactions of the Charles S. Peirce Society 24/1:131–139*, 1988.

[Rya92] M. Ryan. *Ordered Presentation of Theories: Default Reasoning and Belief Revision.* PhD Thesis, Department of Computing, Imperial College London, 1992.

[Rub90] D-H. Ruben. *Explaining Explanation.* Routledge. London and New York. 1990.

[Sab90] M.Ru. Sabre. Peirce's Abductive Argument and the Enthymeme. *Transactions of the Charles S. Peirce Society 26/3:363–372.* 1990.

[Sal71] W.Salmon. *Statistical Explanation and Statistical Relevance.* University of Pittsburgh Press, pp.29-87, 1971.

[Sal77] W.Salmon. 'A Third Dogma of Empiricism'. in Butts and Hintikka (eds)., *Basic Problems in Methodology and Linguistics*, Reidel, Dordrecht. 1977.

[Sal84] W.Salmon *Scientific Explanation and the Casual Structure of the World.* Princeton: Princeton University Press, 1984.

[Sal90] W.Salmon. *Four Decades of Scientific Explanation.* University of Minnesota Press, 1990.

[Sal92] W.Salmon. 'Scientific Explanation', in W. Salmon etal (eds)., *Introduction to the Philosophy of Science.* Prentice Hall, 1992.

[Sal94] W.Salmon. Causality without Counterfactuals, *Philosophy of Science* 61:297-312. 1994.

[Sco71] D. Scott. 'On Engendering an Illusion of Understanding, *Journal of Philosophy* 68, 787–808, 1971.

[Sha91] E. Shapiro. 'Inductive Inference of Theories from Facts', in J. L. Lassez and G. Plotkin (eds), *Computational Logic: Essays in Honor of Alan Robinson.*. Cambridge, Mass. MIT, 1991.

[Sha70] R. Sharp. 'Induction, Abduction, and the Evolution of Science'. *Transactions of the Charles S. Peirce Society 6/1:17–33*. 1970.

[Sho88] Y. Shoham. *Reasoning about Change. Time and Causation from the standpoint of Artificial Intelligence*. The MIT Press, Cambridge, Mass, 1988.

[SL90] J. Shrager and P. Langley (eds). *Computational Models of Scientific Discovery and Theory Formation*. San Meato: Morgan Kaufmann. 1990.

[SLB81] H. Simon, P. Langley, and G. Bradshaw. 'Scientific Reasoning as Problem Solving'. *Synthese* 47 pp. 1–27, 1981.

[SUM96] H. Sipma, T. Uribe, and Z. Manna. 'Deductive Model Checking'. Paper presented at the *Fifth CSLI Workshop on Logic, Language, and Computation*. Center for the Study of Language and Information, Stanford University. June 1996.

[Smu68] R.M. Smullyan. *First Order Logic*. Springer Verlag, 1968.

[Sos75] E. Sosa, (ed). *Causation and Conditionals*. Oxford University Press. 1975.

[Ste83] W. Stegmüller. *Erklärung, Begründung, Kausalität*, second edition, Springer Verlag, Berlin, 1983.

[Sti91] M. Stickel. 'Upside-Down Meta-Interpretation of the Model Elimination Theorem-Proving Procedure for Deduction and Abduction'. Technical Report. Institute for New Generation Computer Technology. (ICOT) TR-664. 1991.

[Sup96] P. Suppes. *Foundations of Probability with Applications*. Cambridge University Press. 1996.

[Tam94] A.M. Tamminga, 'Logics of Rejection: Two Systems of Natural Deduction', in *Logique et analyse*, 37(146), p. 169. 1994.

[Tan92] Yao Hua Tan. *Non-Monotonic Reasoning: Logical Architecture and Philosophical Applications*. PhD Thesis, University of Amsterdam, 1992.

[Tha77] P.R. Thagard. The Unity of Peirce's Theory of Hypothesis. *Transactions of the Charles S. Peirce Society 13/2:112–123*. 1977.

[Tha88] P.R. Thagard. *Computational Philosophy of Science*. Cambridge, MIT Press. Bradford Books. 1988.

[Tha92] P.R. Thagard. *Conceptual Revolutions*. Princeton University Press. 1992.

[Tho81] P. Thompson. 'Bolzano's deducibility and Tarski's Logical Consequence', *History and Philosophy of Logic* 2, 11–20. 1981.

[Tij97] A. ten Teije *Automated Configuration of Problem Solving Methods in Diagnosis*. PhD Thesis, University of Amsterdam, 1997.

[Tou58] S. Toulmin. *The Uses of Argument*. Cambridge University Press, Cambridge, 1958.

[Tro96] A.S. Troelstra. *Basic Proof Theory*. Cambridge University Press. 1996.

[Vis97] H. Visser. *Procedures of Discovery*. Manuscript. Katholieke Universiteit Brabant. Tilburg. The Netherlands. 1997.

[Wil94] M.A. Williams. 'Explanation and Theory Base Transmutations', in *Proceedings of the European Conference on Artificial Intelligence*, ECAI'94, pp. 341–246. 1994.

[Wri63] G.H. von Wright. *The Logic of Preference*. Edinburgh, University Press. 1963.

Titles in the ILLC Dissertation Series:

ILLC DS-93-01: **Paul Dekker**
Transsentential Meditations; Ups and downs in dynamic semantics

ILLC DS-93-02: **Harry Buhrman**
Resource Bounded Reductions

ILLC DS-93-03: **Rineke Verbrugge**
Efficient Metamathematics

ILLC DS-93-04: **Maarten de Rijke**
Extending Modal Logic

ILLC DS-93-05: **Herman Hendriks**
Studied Flexibility

ILLC DS-93-06: **John Tromp**
Aspects of Algorithms and Complexity

ILLC DS-94-01: **Harold Schellinx**
The Noble Art of Linear Decorating

ILLC DS-94-02: **Jan Willem Cornelis Koorn**
Generating Uniform User-Interfaces for Interactive Programming Environments

ILLC DS-94-03: **Nicoline Johanna Drost**
Process Theory and Equation Solving

ILLC DS-94-04: **Jan Jaspars**
Calculi for Constructive Communication, a Study of the Dynamics of Partial States

ILLC DS-94-05: **Arie van Deursen**
Executable Language Definitions, Case Studies and Origin Tracking Techniques

ILLC DS-94-06: **Domenico Zambella**
Chapters on Bounded Arithmetic & on Provability Logic

ILLC DS-94-07: **V. Yu. Shavrukov**
Adventures in Diagonalizable Algebras

ILLC DS-94-08: **Makoto Kanazawa**
Learnable Classes of Categorial Grammars

ILLC DS-94-09: **Wan Fokkink**
Clocks, Trees and Stars in Process Theory

ILLC DS-94-10: **Zhisheng Huang**
Logics for Agents with Bounded Rationality

ILLC DS-95-01: **Jacob Brunekreef**
On Modular Algebraic Protocol Specification

ILLC DS-95-02: **Andreja Prijatelj**
Investigating Bounded Contraction

ILLC DS-95-03: **Maarten Marx**
Algebraic Relativization and Arrow Logic

ILLC DS-95-04: **Dejuan Wang**
Study on the Formal Semantics of Pictures

ILLC DS-95-05: **Frank Tip**
Generation of Program Analysis Tools

ILLC DS-95-06: **Jos van Wamel**
Verification Techniques for Elementary Data Types and Retransmission Protocols

ILLC DS-95-07: **Sandro Etalle**
Transformation and Analysis of (Constraint) Logic Programs

ILLC DS-95-08: **Natasha Kurtonina**
Frames and Labels. A Modal Analysis of Categorial Inference

ILLC DS-95-09: **G.J. Veltink**
Tools for PSF

ILLC DS-95-10: **Giovanna Cepparello**
Studies in Dynamic Logic

ILLC DS-95-11: **W.P.M. Meyer Viol**
Instantial Logic. An Investigation into Reasoning with Instances

ILLC DS-95-12: **Szabolcs Mikulás**
Taming Logics

ILLC DS-95-13: **Marianne Kalsbeek**
Meta-Logics for Logic Programming

ILLC DS-95-14: **Rens Bod**
Enriching Linguistics with Statistics: Performance Models of Natural Language

ILLC DS-95-15: **Marten Trautwein**
Computational Pitfalls in Tractable Grammatical Formalisms

ILLC DS-95-16: **Sophie Fischer**
The Solution Sets of Local Search Problems

ILLC DS-95-17: **Michiel Leezenberg**
Contexts of Metaphor

ILLC DS-95-18: **Willem Groeneveld**
Logical Investigations into Dynamic Semantics

ILLC DS-95-19: **Erik Aarts**
Investigations in Logic, Language and Computation

ILLC DS-95-20: **Natasha Alechina**
Modal Quantifiers

ILLC DS-96-01: **Lex Hendriks**
Computations in Propositional Logic

ILLC DS-96-02: **Angelo Montanari**
Metric and Layered Temporal Logic for Time Granularity

ILLC DS-96-03: **Martin H. van den Berg**
Some Aspects of the Internal Structure of Discourse: the Dynamics of Nominal Anaphora

ILLC DS-96-04: **Jeroen Bruggeman**
Formalizing Organizational Ecology

ILLC DS-97-01: **Ronald Cramer**
Modular Design of Secure yet Practical Cryptographic Protocols

ILLC DS-97-02: **Nataša Rakić**
Common Sense Time and Special Relativity

ILLC DS-97-03: **Arthur Nieuwendijk**
On Logic. Inquiries into the Justification of Deduction