Chapter 3

Example: Inductive Generalization

The insights arrived at in the previous chapter are here applied to a form of *ampliative* reasoning: inductive generalization. There are three aspects: (i) interpreting the data in view of background knowledge, (ii) deriving generalizations from data, and (iii) making further choices, which lead to further generalizations, in terms of one's world view, personal constraints, and other conjecture-generating considerations. These aspects will be discussed separately.

As the considered reasoning forms concern scientific methods, it is desirable that a large number of *variants* are obtained. This will indeed be the case. Which variant is adequate in which context is a different matter, but I shall point to the kind of considerations that may be invoked to settle the matter.

3.1 Promises

On the conventions introduced in Chapter 1, the adaptive logics we have met in Chapter 2 are corrective in that their lower limit logic is weaker than **CL**. In the present chapter I shall introduce logics of inductive generalization, which are ampliative adaptive logics. I shall rely on insights concerning defeasible reasoning that were gained in the previous chapter.

It is often said that there is no logic of induction. This is mistaken: the present chapter contains logics of induction—logics of inductive generalization to be precise. These logics are not a contribution to the great tradition of Carnapian inductive logic—see for example [Kui00, Ch. 4]. They are logics of induction in the most straightforward sense of the term, logics that from a set of empirical data and possibly a set of background generalizations lead to a set of generalizations and their consequences.

That there will be many logics of induction is as expected. Indeed, these logics are formally stringent formulations of inductive methods. Also, the consequences derivable by these logics may be given several interpretations. One interpretation would be that the derived generalizations and predictions are provisionally accepted as true. On a different interpretation, the generalizations are selected for further testing in the sense of Popper.

The logics will have a proof theory and a semantics. They are characterized in a formally decent way, their metatheory may be phrased in precise terms, as will be shown in Chapters 4 and 5, and, most importantly, they aim at explicating people's actual reasoning that leads to inductive generalizations.

Two problems will be tackled in this chapter. The first concerns the transition from a set of empirical data to inductive generalizations. This problem is solved by the logics of inductive generalization properly. But there is obviously a need, in order to make the enterprise realistic, to combine data with background knowledge before moving to generalizations. I shall show that this different problem may also be solved by means of adaptive logics.

All logics will be formulated in $\mathcal{L}_s$, the standard predicative language. Consequences follow either deductively or inductively from the premises—I shall take them to follow deductively if they follow from the premises by Classical Logic. The main interest of the logics obviously concerns the inductive consequences. These comprise, first and foremost, empirical generalizations. They also comprise deductive consequences of the premises and of the generalizations. These consequences include singular statements that may serve the purposes of prediction and explanation.[1]

Adopting the standard predicative language is obviously a restriction. Actually, the generalizations will even be more restricted. I shall restrict the attention to unary predicates and to generalizations that do not refer to singular objects. The first restriction is introduced to avoid the complications involved in relational predicates. First problems should be solved first. The second restriction will be explained later in this chapter. The restrictions rule out statistical generalizations as well as quantitative predicates (lengths, weights, etc.). In other words, I shall stick to the basic case, leaving the removal of some restrictions for future research.

The logics presented will not take into account degrees of confirmation or the number of confirming and disconfirming instances. I also shall disregard such problems as discovery and creativity as well as conceptual change and other forms of conceptual dynamics. Moreover, I shall disregard inconsistent background theories.[2] So the logics presented are only a starting point.

When working on inductive generalization, for example on [Bat05a] and [BH01], I wondered why systems as simple and clarifying as the logics articulated below had not been presented a long time ago.[3] The reason is presumably that their formulation presupposes some familiarity with the adaptive logic programme. Yet, the logics are extremely simple and extremely promising.

Sections 3.2 and 3.3 are based on [Bat05a], part of Section 3.4 on [BH01].[4] Other materials in this chapter result from research that was started by me and

[1] A potential explanation in terms of a theory T may be established by means of a prediction that derives from generalizations in a different domain.

[2] These may be approached by replacing **CL** in the logics by **CLuN** and combining the ampliative, inductive, aspect with a corrective, inconsistency-handling, one.

[3] In the form of a formal logic, that is. There are connections with Mill's canons. There are also connections with Reichenbach's straight rule, if restricted to general hypotheses, and with Popper's conjectures and refutations. Articulating the formal logic drastically increases precision, as we shall see.

[4] The reader familiar with those papers will note certain differences in the formal presentation. This may be a nuisance, but it is unavoidable that a better presentation is reached only after a number of years.

in which Mathieu Beirlaen and Frederik Van De Putte joined in later. Pieter Robberechts wrote a set of computer programs that offered 'empirical' help and thus accelerated theoretic insights.

3.2 A First Logic of Inductive Generalization

Children have a heavy tendency to generalize, which has a clear survival value. Of course, all simple empirical generalizations are false—compare [Pop73, p. 10]. Our present scientific (and other) knowledge results from methods that are more sophisticated than the aforementioned tendency. The methodological improvement was learned from experience and free inquiry. It does not follow, however, that our knowledge is the result of an urge that is qualitatively different from children's tendency to generalize, nor that it would be the outcome of forms of reasoning that are qualitatively different from children's. The improvement is related to performing systematic research and systematizing knowledge.

Consider the case in which only a set of empirical data is available. Where these are our only premises, what shall we want to derive from them? Apart from the **CL**-consequences of the premises, we want to derive some general hypotheses. Only in doing so may we hope to get a grasp on the world—to understand the world and to act in it. Moreover, from our premises and hypotheses together we shall want to derive **CL**-consequences to test the hypotheses, to predict facts, and to explain facts. Is there a consequence relation that provides us with all this? Well, let us see.

Generalizations that are inductively derived from a set of data should be compatible with this set. They should moreover be *jointly* compatible with it.[5] The logic of compatibility—see Section 9.2—provides us with the set of all statements compatible with Γ. However, to select a set of jointly compatible statements *in a justified way* seems hopeless. For any statement A that does not deductively follow from the premises, there is a set of statements Δ such that the members of Δ are jointly compatible with the premises whereas the members of $\Delta \cup \{A\}$ are not. However, it turns out possible to use joint compatibility as a criterion provided one considers only generalizations as restricted in Section 3.1.

Mere compatibility is not difficult to grasp. A generalization is compatible with a set of data just in case the data do not falsify it. This means, just in case its negation is not derivable from the data. Let us try out that crude idea, which agrees with the hypothetico-deductive approach.

As in the previous chapter, there will be an unconditional rule, RU, and a conditional rule RC. In this chapter, the unconditional rule will take care of **CL**-consequences and the conditional rule will handle the ampliative consequences. The conditional rule will allow one to introduce a generalization on a condition, which will be (the singleton comprising) the negation of the generalization. So where the abnormalities in the previous chapter were inconsistencies, they will now be negations of generalizations. To keep things simple, let the strategy be

[5] For the present context, a formula A is compatible with a set of formulas Γ iff $\Gamma \cup \{A\}$ is consistent; alternatively iff $\Gamma \nvdash_{\mathbf{CL}} \neg A$. The members of a finite or infinite set Δ are jointly compatible with Γ iff $\Gamma \cup \Delta$ is consistent; alternatively iff there are no $A_1, \ldots, A_n \in \Delta$ such that $\Gamma \nvdash_{\mathbf{CL}} \neg A_1 \vee \ldots \vee \neg A_n$. Note that, on this convention, nothing is compatible with an inconsistent Γ.

Reliability, which means that a line is marked at a stage iff its condition overlaps with the set of unreliable formulas. Remember that the formula of a line that is marked at a stage is considered as not derived at that stage.

Why are negations of generalizations called abnormalities? There is on the one hand a technical justification for doing so: they serve the purpose that was served by contradictions in the previous chapter. There is also a philosophical justification: in the present context, negations of generalizations are considered as false until and unless proven otherwise, viz. until and unless the generalization is falsified. Induction is connected to the presupposition that the world is as uniform as allowed by the data—the connection between induction and uniformity is referred to in [Car52] and elsewhere. The negation of a generalization expresses a lack of uniformity and so is considered an abnormality.

Let us call this logic of induction $\mathbf{LI}^r$, the superscripted r referring to Reliability. Here is the start of a very simple proof from $\Gamma_1 = \{(Pa \wedge Pb) \wedge Pc, Rb \vee \neg Qb, Rb \supset \neg Pb, (Sa \wedge Sb) \wedge Qa\}$. Some **CL**-consequences are derived and two generalizations are introduced.

1	$(Pa \wedge Pb) \wedge Pc$	premise	$\emptyset$
2	$Rb \vee \neg Qb$	premise	$\emptyset$
3	$Rb \supset \neg Pb$	premise	$\emptyset$
4	$(Sa \wedge Sb) \wedge Qa$	premise	$\emptyset$
5	Pa	1; RU	$\emptyset$
6	Pb	1; RU	$\emptyset$
7	Qa	4; RU	$\emptyset$
8	Sa	4; RU	$\emptyset$
9	Sb	4; RU	$\emptyset$
10	$\forall x(Px \supset Sx)$	RC	$\{\neg\forall x(Px \supset Sx)\}$
11	$\forall x(Px \supset Qx)$	RC	$\{\neg\forall x(Px \supset Qx)\}$

The two generalizations are considered as 'conditionally' true, as true until falsified. The sole member of $\{\neg\forall x(Px \supset Sx)\}$ has to be false in order for $\forall x(Px \supset Sx)$ to be derivable and similarly for line 11. Conditionally derived formulas may obviously be combined by RU. As expected, the condition of the derived formula is the union of the conditions of the formulas from which it is derived. Here is an example:

12	$\forall x(Px \supset (Qx \wedge Sx))$	10, 11; RU	$\{\neg\forall x(Px \supset Sx), \neg\forall x(Px \supset Qx)\}$

The interpretation of the condition of line 12 is obviously that $\forall x(Px \supset (Qx \wedge Sx))$ should be considered as not derived if $\neg\forall x(Px \supset Sx)$ or $\neg\forall x(Px \supset Qx)$ turns out to be true. Actually, it is not difficult to see that $\neg\forall x(Px \supset Qx)$ is indeed derivable from the premises.

13	$\neg Rb$	3, 6; RU	$\emptyset$
14	$\neg Qb$	2, 13; RU	$\emptyset$
15	$\neg\forall x(Px \supset Qx)$	6, 14; RU	$\emptyset$

So lines 11 and 12 have to be *marked* and their formulas are considered as not inductively derived from the premises. I repeat lines 10–15 with the correct marks at stage 15:

10	$\forall x(Px \supset Sx)$	RC	$\{\neg\forall x(Px \supset Sx)\}$	
11	$\forall x(Px \supset Qx)$	RC	$\{\neg\forall x(Px \supset Qx)\}$	$\checkmark^{15}$
12	$\forall x(Px \supset (Qx \wedge Sx))$	10, 11; RU	$\{\neg\forall x(Px \supset Sx), \neg\forall x(Px \supset Qx)\}$	$\checkmark^{15}$
13	$\neg Rb$	3, 6; RU	$\emptyset$	
14	$\neg Qb$	2, 13; RU	$\emptyset$	
15	$\neg\forall x(Px \supset Qx)$	6, 14; RU	$\emptyset$	

Derived generalizations obviously entail predictions by RU, whence these receive the same condition as the generalizations on which they rely.

16	Pc	1; RU	$\emptyset$
17	Sc	16, 10; RU	$\{\neg\forall x(Px \supset Sx)\}$

At this point I have to report about a fascinating phenomenon. Although we started from a simple hypothetico-deductive approach, based on falsification,[6] it turns out that some abnormalities are connected. Not only falsification, but also the connection between abnormalities, which comes to joint falsification,[7] will prevent certain generalizations from being adaptively derivable. A nice example is obtained if one attempts to derive that all objects have a certain property. In the derived atoms (primitive formulas and their negations) we have objects known to be P, but none known to be $\neg P$, and we have objects known to be $\neg R$, but none known to be R. So it seems attractive to introduce two generalizations expressing this. The first generalization, $\forall x Px$, is indeed finally derivable, but the second is not.[8] I shall first present the extension of the proof and then comment upon it.

18	$\forall x Px$	RC	$\{\neg\forall x Px\}$	
19	$\forall x\neg Rx$	RC	$\{\neg\forall x\neg Rx\}$	$\checkmark^{22}$
20	$Ra \vee \neg Ra$	RU	$\emptyset$	
21	$Ra \vee (Qa \wedge \neg Ra)$	20, 7; RU	$\emptyset$	
22	$\neg\forall x\neg Rx \vee \neg\forall x(Qx \supset Rx)$	21; RU	$\emptyset$	

The formula unconditionally derived at line 22 is a *Dab*-formula, a disjunction of abnormalities. It is moreover a minimal *Dab*-formula at stage 22 (and actually also a minimal *Dab*-consequence of the premise set).[9] So line 19 is marked at this stage and is actually marked in every extension of the stage. The generalization $\forall x Px$, derived at line 18, is finally derived. There is no minimal *Dab*-consequence of the premises of which $\neg\forall x Px$ is a disjunct.

I claimed I was reporting on a fascinating phenomenon. Indeed, the adaptive approach reveals that connected abnormalities may prevent one from deriving the related generalizations. So, although we started from a simple (and possibly simplistic) hypothetico-deductive approach that deems a generalization non-derivable in case it is falsified, we arrived at the insight that a generalization may not be derivable because it belongs to a minimal (finite) set of generalizations one of which is bound to be falsified by the data. Some of the literature verborgen
simply missed this point. Time and again one finds claims that induction is too

[6] On present conventions, a set of data Γ falsifies A iff $\Gamma \vdash_{\mathbf{CL}} \neg A$.

[7] Two generalizations are jointly falsified by a set of data iff their conjunction is falsified by the set of data.

[8] Previous examples of generalizations were all universally quantified implications. I never required that this be the case. Moreover, $\forall x\neg Rx$ is **CL**-equivalent to, for example, $\forall x((Px \vee$

permissive a method in that it leads to the acceptance of generalizations that are jointly incompatible with the data. So all inductive generalization would involve an arbitrary choice. The adaptive approach takes this feature into account from the very start, as lines 18–22 show, and thus avoids an arbitrary choice.

This deserves a further comment. Apparently joint compatibility is not seen as a suitable means to distinguish between generalizations that may sensibly be derived from a set of data and those that may not sensibly be so derived. The reason is obviously that every generalization belongs to a set of generalizations that is incompatible with the data. However, $\mathbf{LI}^r$ solves this predicament. It allows one to derive the generalizations that do *not* belong to a *minimal* set of generalizations that are jointly incompatible with the data. For this reason $\forall x \neg Rx$ is not derivable from the considered data, but $\forall x Px$ is. Indeed, $\neg\forall x Px$ is not a disjunct of a minimal *Dab*-consequence of the data—note that a *Dab*-formula is equivalent to the negation of a conjunction of generalizations. It is precisely by invoking *minimality* that the above criterion is made to do its job.[10] So, as I suggested, the adaptive approach gets it right from the beginning.

That connected abnormalities play an essential role may also be seen by considering a predicate that does not occur in the premises. T is such a predicate, and one might introduce $\forall x(Qx \supset Tx)$ to see what becomes of it. Not much, as the following extension of the proof shows.

23	$\forall x(Qx \supset Tx)$	RC	$\{\neg\forall x(Qx \supset Tx)\}$	$\checkmark^{25}$
24	$\forall x(Qx \supset \neg Tx)$	RC	$\{\neg\forall x(Qx \supset \neg Tx)\}$	$\checkmark^{25}$
25	$\neg\forall x(Qx \supset Tx) \vee \neg\forall x(Qx \supset \neg Tx)$	7; RU	$\emptyset$	

Obviously, Γ_1 does not contradict $\forall x(Qx \supset Tx)$. However, it contradicts $\forall x(Qx \supset Tx) \wedge \forall x(Qx \supset \neg Tx)$, which is noted on line 25 and causes lines 23 and 24 to be marked. Suppose that we used a marking definition that causes a line to be marked only if an element of its condition is *derived* unconditionally. On such a definition, lines 23 and 24 would both be unmarked and so would be a line at which one would derive $Ta \wedge \neg Ta$ from 7, 23, and 24. So triviality would result. Incidentally, the same reasoning applies to the marks caused by line 22.

Allow me to insert a short intermission at this point. By applying RU, viz. Addition, one may derive

$$\forall x(Qx \supset Tx) \vee \forall x(Qx \supset \neg Tx)$$

on two different conditions: from line 23 on the condition $\{\neg\forall x(Qx \supset Tx)\}$ and from line 24 on the condition $\{\neg\forall x(Qx \supset \neg Tx)\}$. The two lines on which the disjunction would be so derived would still be marked on the Reliability strategy. However, once both occur, they are unmarked if Minimal Abnormality is the strategy, so for the logic $\mathbf{LI}^m$. This illustrates that Minimal Abnormality

$\neg Px) \supset \neg Rx)$.

[9]Remember that every *Dab*-formula entails infinitely many *Dab*-formulas by Addition. That does not make the added disjuncts unreliable because only the minimal *Dab*-formulas are taken into account for defining $U(\Gamma)$—see the previous chapter.

[10]The compactness of **CL** warrants that every *minimal* set of formulas (any formulas that is), the members of which are jointly incompatible with a premise set (any premise set), is finite.

leads to some more consequences than Reliability, just as it did in the previous chapter.

Two further things have to be cleared up. First, which are precisely the restrictions on the generalizations? Actually, these restrictions are introduced as restrictions on the abnormalities, which comes to the same in the context of $\mathbf{LI}^r$. Let me first illustrate the need for a restriction.

Suppose that it were allowed to introduce the generalization $\forall x((Qx \vee \neg Qx) \supset \neg Sc)$ on the condition $\{\neg\forall x((Qx \vee \neg Qx) \supset \neg Sc)\}$ in the preceding proof. From line 1 follows

$$\neg\forall x(Px \supset Sx) \vee \neg\forall x((Qx \vee \neg Qx) \supset \neg Sc)$$

by RU. This would cause the line on which $\forall x((Qx \vee \neg Qx) \supset \neg Sc)$ is derived to be marked. However, it would also cause line 10 to be marked. In this way, every conditional line could be marked because, for every generalization, there is a formula such that both are jointly incompatible with the premises—note that $\forall x((Qx \vee \neg Qx) \supset \neg Sc)$ is **CL**-equivalent to $\neg Sc$. Similar troubles arise if it were allowed to introduce such hypotheses as $\forall x((Qx \vee \neg Qx) \supset (\exists y)(Py \wedge \neg Sy))$.

The previous paragraph starts off with a supposition. Indeed, the restrictions introduced in Section 3.1 prevent such troubles. Let $\mathcal{F}_s^{f1}$ denote the purely functional formulas of rank 1: formulas in which occur no sentential letters, no individual constants, no quantifiers, and no predicates of a rank higher than 1. Define an *abnormality* (in the present context) as a formula of the form $\neg\forall A$, in which $\forall$ abbreviates a universal quantifier over every variable free in A and $A \in \mathcal{F}_s^{f1}$. This will obviously have a dramatic effect on the generalizations that can be derived.[11] Some people will raise a historical objection to this restriction. Kepler's laws explicitly refer to the sun, and Galilei's law of free fall refers to the earth. This, however, is related to the fact that the earth, the sun, and the moon had a specific status in the Ptolemaic world-view, which they were losing very slowly in the days of Kepler and Galilei. In the Ptolemaic world-view, each of those objects was taken, just like God, to be the only object of a specific kind. So the generalizations referred to kinds of objects, rather than to specific objects—by Newton's time, any possible doubt about this had been removed.[12]

The other point to be cleared up concerns the dynamic character of the proofs. Some will reason as follows. The set of data consists of singular statements and is forcibly finite. Which *Dab*-formulas are derivable from it is decidable in view of the restriction on abnormalities. So one may avoid, and hence should avoid, applications of the conditional rule that are later marked.

I strongly disagree with the normative conclusion. In the context of the present logic, marked lines may indeed be avoided. But why should they be avoided? For one thing, the proposed alternative is not static proofs, but a complication of dynamic proofs—these are studied in Section 4.7. Deriving all *Dab*-consequences of a set of singular premises requires long and inefficient proofs—compare with Chapter 10. This holds true even if abnormalities are re-

[11] It is possible to devise a formal language in which the effect of the restriction is reduced to nil. This is immaterial because such languages are not used in the empirical sciences, to which we want to apply the present logic. But indeed, the formal restriction hides one on content: all predicates should be well entrenched, and not abbreviate, for example, identity to an individual constant.

[12] Even in the Ptolemaic era, those objects were identified in terms of well entrenched properties—properties relating to their kind, not to accidental qualities. The non-physical example is even more clear: God had *no* accidental properties.

stricted to one element out of every set of equivalent abnormalities—the number of these abnormalities is finite, whereas the set of all derivable abnormalities is not. Even if the *Dab*-consequences of Γ are so restricted, the derivability of a generalization is not established by the proof itself, but only by a reasoning in the metalanguage—compare Section 4.4. So the alternative is not static proofs. Given that we need dynamic proofs for other logics (see Corollary 5.8.4) and given that dynamic proofs are well studied and that procedures for establishing final derivability have been developed, dynamic proofs for inductive generalization are more attractive than their alternative. And there is more.

The logic I am trying to articulate here is a very restricted one. An obvious example of a restriction is that it does not take background knowledge into account. As soon as some of the restrictions are eliminated, we obtain a consequence relation for which there is no positive test and which will require dynamic proofs. There is a further argument and it should not be taken lightly. Adaptive logics aim at explicating actual inductive reasoning and this is dynamic. Given our finite brains, it would be a bad policy to make inductive hypotheses contingent on complete procedural certainty. To do so would slow down our thinking, often paralyse it. This does not mean that one may neglect criteria for final derivability. It only means that one often bases decisions on incomplete knowledge—see Section 4.10 for a formal approach to the analysis of deductive information.

In Section 1.3, I stated that adaptive logics form a qualitative approach. It seems wise to recall this here and to point out one of its consequences. The logic $\mathbf{LI}^r$ does not take into account how many instances of generalizations occur in the data. So a premise set like $\Gamma_2 = \{Pa, \neg Pb \vee Qb\}$ may be seen as expressing: all we know is that some objects have property P and others have not property P or have property Q. In other words, extending Γ_2 with $\{Pc, Pd, \neg Pe \vee Qe\}$ does not affect the consequence set as far as generalizations are concerned—it obviously affects the singular consequences.

It is time to present a precise formulation of our first inductive logic, $\mathbf{LI}^r$. A little reflection on derivable rules readily reveals that the proofs are governed by rules that look exactly like the ones from the previous chapter, except that **CLuN** has to be replaced by **CL**.

ev. tabular splitsen vr paginering

Prem If $A \in \Gamma$:

$$\begin{array}{ll} \ldots & \ldots \\ \hline A & \emptyset \end{array}$$

RU If $A_1, \ldots, A_n \vdash_{\mathbf{CL}} B$:

$$\begin{array}{ll} A_1 & \Delta_1 \\ \ldots & \ldots \\ A_n & \Delta_n \\ \hline B & \Delta_1 \cup \ldots \cup \Delta_n \end{array}$$

RC If $A_1, \ldots, A_n \vdash_{\mathbf{CL}} B \vee Dab(\Theta)$:

$$\begin{array}{ll} A_1 & \Delta_1 \\ \ldots & \ldots \\ A_n & \Delta_n \\ \hline B & \Delta_1 \cup \ldots \cup \Delta_n \cup \Theta \end{array}$$

Let me show that these rules are adequate. In the example proof, I used RC to introduce a generalization A on the condition $\{\neg A\}$. The rule RC as formulated here obviously enables one to do so. For any generalization A, $A \vee$

$\neg A$ is derivable by RU from any Γ and the present RC enables one to derive A on the condition $\neg A$ from $A \vee \neg A$. The present rule RC seems to enable one to derive more than that, but actually it does not. Suppose that $\Gamma \vdash_{\mathbf{CL}} B \vee (\neg A_1 \vee \ldots \vee \neg A_n)$ in which all $\neg A_i$ are abnormalities. It follows that $\Gamma \vdash_{\mathbf{CL}} (A_1 \wedge \ldots \wedge A_n) \supset B$. So one may extend any proof from Γ by the following lines, in which RC is only used as in the preceding example proof.

j_0	$(A_1 \wedge \ldots \wedge A_n) \supset B$	RU	$\emptyset$
j_1	A_1	RC	$\{\neg A_1\}$
$\vdots$			
j_n	A_n	RC	$\{\neg A_n\}$
j_{n+1}	B	$j_0, \ldots, j_n$; RU	$\{\neg A_1, \ldots, \neg A_n\}$

In other words, the rule RC as used in the example proof enables one to derive the general rule RC.

Where one defines $U_s(\Gamma)$ as in the previous chapter, but adjusted to the new abnormalities, marking is governed by Definition 2.3.2 and final derivability follows Section 2.3.5.

Formulating the $\mathbf{LI}^r$-semantics is easy. The minimal *Dab*-consequences of Γ are defined as in the previous chapter (from the present abnormalities) and from them one defines $U(\Gamma)$. Define $Ab(M)$ as the set of abnormalities verified by M. A **CL**-model M of Γ is a reliable model of Γ iff $Ab(M) \subseteq U(\Gamma)$. $\Gamma \vDash_{\mathbf{LI}^r} A$ iff A is verified by all reliable models of Γ.

Some paragraphs ago, I mentioned the minimal Abnormality strategy, which defines the logic $\mathbf{LI}^m$. Its marking definition is wholly analogous to Definition 2.3.3 and its semantics to the semantics of $\mathbf{CLuN}^m$.[13] We shall see later that $\mathbf{LI}^r$ and $\mathbf{LI}^m$ have some awkward properties. First however, we move to a different matter.

3.3 Heuristic Matters and Further Comments

Some people think that all adaptive reasoning (including all non-monotonic reasoning) should be explicated in terms of heuristic moves rather than in terms of logic properly. For their instruction and confusion, some basics of the heuristics of the adaptive logic **LI** deserve to be spelled out. There is another reason to insert the present section. $\mathbf{LI}^r$ and $\mathbf{LI}^m$ are extremely simple logics of inductive generalization. Nevertheless, they clearly are capable of guiding research. I shall discuss some elementary considerations in this respect.

One might think that $\mathbf{LI}^r$ allows one to derive anything that concerns a predicate about which one has no information at all. However, the 'dependency' between abnormalities as shown in the table below prevents one from doing so. Suppose that one is interested in the relation between P and Q and adds the following line to a proof.

i	$\forall x(Px \supset Qx)$	RC	$\{\neg\forall x(Px \supset Qx)\}$

As (3.1) is a **CL**-theorem, it may be derived in the proof and causes line i to be L-marked.

$$\neg\forall x(Px \supset Qx) \vee \neg\forall x(Px \supset \neg Qx) \vee \neg\forall x(\neg Px \supset Qx) \vee \neg\forall x(\neg Px \supset \neg Qx) \quad (3.1)$$

[13]Replacing the lower limit logic and the set of abnormalities is obvious.

In order to prevent $\forall x(Px \supset Qx)$ from being marked, one needs to unconditionally derive a 'sub-disjunction' of (3.1) that does not contain $\neg\forall x(Px \supset Qx)$. How does one do so? Here is a little instructive table.

an instance of	enables one to derive
Px	$\neg\forall x(Px \supset Qx) \vee \neg\forall x(Px \supset \neg Qx)$
$\neg Px$	$\neg\forall x(\neg Px \supset Qx) \vee \neg\forall x(\neg Px \supset \neg Qx)$
Qx	$\neg\forall x(Px \supset \neg Qx) \vee \neg\forall x(\neg Px \supset \neg Qx)$
$\neg Qx$	$\neg\forall x(Px \supset Qx) \vee \neg\forall x(\neg Px \supset Qx)$
$Px \wedge Qx$	$\neg\forall x(Px \supset \neg Qx)$
$Px \wedge \neg Qx$	$\neg\forall x(Px \supset Qx)$
$\neg Px \wedge Qx$	$\neg\forall x(\neg Px \supset \neg Qx)$
$\neg Px \wedge \neg Qx$	$\neg\forall x(\neg Px \supset Qx)$

The table clearly suggests which kind of information one should gather in specific circumstances. If $\neg\forall x(Px \supset Qx) \vee \neg\forall x(Px \supset \neg Qx)$ is a minimal *Dab*-formula, it will cause line i to be marked. In order to prevent this, one needs an instance of $Px \wedge Qx$ and one needs to derive $\neg\forall x(Px \supset \neg Qx)$ from it. If $\neg\forall x(Px \supset Qx) \vee \neg\forall x(\neg Px \supset Qx)$ is a minimal *Dab*-formula, line i will also be marked. To prevent this, one needs an instance of $\neg Px \wedge \neg Qx$ and one needs to derive $\neg\forall x(\neg Px \supset Qx)$ from it.

It is worth noting that conditional derivation does not in any way confuse the relations spelled out in the table. Suppose that one derives a generalization by RC, and next derives a prediction from the data and these generalization by RU. In this way one may very well be able to derive, by RU, a disjunction of abnormalities. However, the prediction as well as the disjunction of abnormalities will be derived on the condition on which the generalization was derived. So the disjunction of abnormalities does not count as a *Dab*-formula and does not have any effect on the marks. In simple adaptive logics—combined adaptive logics will be introduced later in this chapter—a *Dab*-formula counts only as a minimal *Dab*-formula iff it is derived on the empty condition. As I shall prove (Lemma 4.4.1), a formula A derived on a condition Δ corresponds to the unconditional derivation of $A \vee Dab(\Delta)$. So if A is a disjunction of abnormalities, the derivable *Dab*-formula that has effects on the marks is $A \vee Dab(\Delta)$. Returning to the point I was making: extending the proof with applications of RC will never enable one to have a generalization unmarked which otherwise would be marked. This is provable: see Theorem 5.6.2.

As conditional derivation causes no trouble, let us return to the heuristic guidance provided by **LI**. The table seems to make things extremely simple, but there is a fascinating complication. Suppose that $\Gamma_3 = \{Pa, Ra\}$, that someone writes the proof 1-3 and, relying on the table, adds 4.

1	Pa	premise	$\emptyset$	
2	Ra	premise	$\emptyset$	
3	$\forall x(Px \supset Qx)$	RC	$\{\neg\forall x(Px \supset Qx)\}$	$\checkmark^4$
4	$\neg\forall x(Px \supset Qx) \vee \neg\forall x(Px \supset \neg Qx)$	1; RU	$\emptyset$	

In order that line 3 be unmarked, the table instructs us to find an instance of $Px \wedge Qx$ and to derive $\neg\forall x(Px \supset \neg Qx)$ from it. Suppose then that such an instance is obtained and is introduced as a new premise (see 5 below). This need not be the end of the story, as the following continuation of the proof illustrates.

3	$\forall x(Px \supset Qx)$	RC	$\{\neg\forall x(Px \supset Qx)\}$	$\checkmark^{7}$
4	$\neg\forall x(Px \supset Qx) \vee \neg\forall x(Px \supset \neg Qx)$	1; RU	$\emptyset$	
5	$Pb \wedge Qb$	premise	$\emptyset$	
6	$\neg\forall x(Px \supset \neg Qx)$	5; RU	$\emptyset$	
7	$\neg\forall x(Px \supset Qx) \vee \neg\forall x(Rx \supset \neg Qx)$	1, 2; RU	$\emptyset$	

So line 3 is still marked and in order to have it unmarked, one needs an instance of $Rx \wedge Qx$. Control is provided by the following simple and intuitive fact:

(†) If the introduction of a generalization G_1 contextually entails a falsifying instance of another generalization G_2, and no falsifying instance of the latter is derivable from the data alone, then $\neg G_1 \vee \neg G_2$ is unconditionally derivable.

Let us apply this to the present example. From $\forall x(Px \supset Qx)$ together with 1 follows $Ra \wedge Qa$, which is a falsifying instance of $\forall x(Rx \supset \neg Qx)$. Obviously, $Ra \wedge Qa$ is derivable on the condition $\{\neg\forall x(Px \supset Qx)\}$. However, $\forall x(Rx \supset \neg Qx)$ is derivable on the condition $\{\neg\forall x(Rx \supset \neg Qx)\}$ and no falsifying instance, viz. instance of $Rx \wedge Qx$, is derivable from the data alone. So $\neg\forall x(Px \supset Qx) \vee \neg\forall x(Rx \supset \neg Qx)$ is derivable on the empty condition. The reasoning behind (†) is common to all adaptive logics. Here is its general form: if A is derivable on the condition Δ, B is derivable on the condition Θ, and A and B exclude each other,[14] then $Dab(\Delta \cup \Theta)$ is unconditionally derivable. This is simply an application of $A \vee C, B \vee D, \neg(A \wedge B) \vdash_{\mathbf{CL}} C \vee D$. In Section 4.4, I shall introduce a handy derived rule, RD, that relies on a simplified form of this reasoning.

One might conclude from (†) that $\mathbf{LI}^r$ advises one to try to falsify only hypotheses that are rivals for the ones one likes to derive. This conclusion is mistaken. The reason to apply logics of inductive generalization is to arrive at reliable knowledge about the world. One question is whether a generalization is derivable from the available data in the sense that it does not belong to a minimal set of generalizations that is jointly incompatible with the data. A very different question is whether the data set was selected in the best possible way.

It is worth mentioning a heuristic aspect that is not related to gathering new information but to deriving the relevant *Dab*-formulas from a set of data. As said in the next to last paragraph, if A is derivable on the condition Δ, B is derivable on the condition Θ, and A and B exclude each other, then $Dab(\Delta \cup \Theta)$ is unconditionally derivable. So by introducing generalizations and by deriving predictions from them, relevant *Dab*-formulas are obtained as it were for free. There are further similar mechanisms. One of them was exemplified before (and is warranted by Lemma 4.4.1): if a formula $Dab(\Delta)$ is derived on the condition Δ', then $Dab(\Delta \cup \Delta')$ is unconditionally derivable. Similarly, if $A(x), B(x) \in \mathcal{F}_s^{f1}$, $A(a)$ is derived on the condition Δ and $B(a)$ is derived on the condition Δ', then the *Dab*-formula $\neg\forall x(A(x) \supset \neg B(x)) \vee Dab(\Delta \cup \Delta')$ is unconditionally derivable—either or both of Δ and Δ' may be empty. So in order to speed up the dynamic proof, it is heuristically advisable (i) to conditionally derive singular statements that instantiate formulas of which no instance was unconditionally

[14] In the present context, it is obvious that this means that they exclude each other in the sense of **CL**. This also holds in general: what is required is that $\{A, B\}$ is **CL**-trivial.

derived and (ii) to derive *Dab*-formulas that change the minimal *Dab*-formulas in comparison to the previous stage of the proof.

Some of the discussed heuristic aspects concern the gathering of new data, in other words premises. Note that these data may be simply added as premises to an ongoing proof. This move will always result in a correct proof (or rather proof stage). So the dynamic proofs are *robust* in this sense.

more comments on expanding premise sets

3.4 Two Alternatives

The abnormalities of $\mathbf{LI}^r$ and $\mathbf{LI}^m$ are formulas of the form $\neg\forall A$ in which $A \in \mathcal{F}_s^{f1}$. I shall first give a different and simpler presentation of these abnormalities.

Let $\mathcal{A}$, the set of *atoms*, be the set comprising all primitive formulas as well as their negations. Note that this set also includes p, $\neg p$, ..., $a = a$, $a = b$, $\neg\, a = a$, $\neg\, a = b$, etc.[15] The set of functional atoms, $\mathcal{A}^f$, comprises the atoms that do not contain sentential letters or individual constants. The set of functional atoms of rank 1, $\mathcal{A}^{f1}$, comprises the functional atoms in which the predicate is of rank 1—so no identities because identity is a 'logical' predicate of rank 2. I now redefine the set of abnormalities of $\mathbf{LI}^r$ and $\mathbf{LI}^m$ as: $\{\exists\neg(A_0 \vee \ldots \vee A_n) \mid A_0, \ldots, A_n \in \mathcal{A}^{f1}; n \geq 0\}$.

It is easily seen that this redefinition leads to a different formulation of the same logics. If $A \in \mathcal{F}_s^{f1}$, then $\forall A$ is **CL**-equivalent to a conjunction of formulas of the form $\forall(B_0 \vee \ldots \vee B_n)$ in which $B_0, \ldots, B_n \in \mathcal{A}^{f1}$. For example, $\forall x(((Px \wedge \neg Qx) \vee Rx) \supset Sx)$ is **CL**-equivalent to $\forall x(\neg Px \vee Qx \vee Sx) \wedge \forall x(\neg Rx \vee Sx)$ and its negation, $\neg\forall x(((Px \wedge \neg Qx) \vee Rx) \supset Sx)$, is **CL**-equivalent to the disjunction of the new abnormalities: $\exists x\neg(\neg Px \vee Qx \vee Sx) \vee \exists x\neg(\neg Rx \vee Sx)$. So the new formulation allows one to introduce $\forall x(((Px \wedge \neg Qx) \vee Rx) \supset Sx)$ on the condition $\{\exists x\neg(\neg Px \vee Qx \vee Sx), \exists x\neg(\neg Rx \vee Sx)\}$ just in case the old formulation allows one to introduce it on the condition $\{\neg\forall x(((Px \wedge \neg Qx) \vee Rx) \supset Sx)\}$. Moreover, if it is possible to derive a *Dab*-formula of the new sort that has an element of the new condition as a disjunct, then it is possible to derive a *Dab*-formula of the old sort that has the (sole) element of the old condition as a disjunct, and *vice versa*. So the line at which the generalization is introduced will be marked in just the same cases.

Redefined or not, $\mathbf{LI}^r$ and $\mathbf{LI}^m$ have an odd property: no model is normal. In the previous chapters, I mentioned that an adaptive logic has a lower limit logic and an upper limit logic. In the present chapter, this lower limit logic is **CL**. The upper limit logic is the logic that does not tolerate abnormalities: all premise sets that entail abnormalities have the trivial consequence set. So the upper limit logic of $\mathbf{LI}^r$ is the trivial logic **Tr**, according to which every formula is derivable from every premise set—for all $\Gamma \subseteq \mathcal{W}_s$, $Cn^{\mathcal{L}_s}_{\mathbf{Tr}}(\Gamma) = \mathcal{W}_s$.

One may think that this odd property results from the fact that the lower limit logic is **CL**. This is mistaken. The property results from the specific interpretation of *uniformity* that underlies $\mathbf{LI}^r$. In Section 3.2, I pointed out that $\mathbf{LI}^r$ presupposes that the world is as uniform as the data permit. If uniformity is identified with the truth of *every* generalization, as in $\mathbf{LI}^r$, no possible world is completely uniform. But uniformity may be interpreted differently: a completely uniform world is one in which all objects have the same properties. Put differently, if something has a property, then everything has this property. So

[15] So some atoms are **CL**-inconsistent, for example $\neg\, a = a$.

$\exists xPx \supset \forall xPx$ and, in general and in terms of abbreviations introduced before, $\exists A \supset \forall A$ for all $A \in \mathcal{F}_s^{f1}$. Some worlds are completely uniform in this sense, although not our world. Fortunately so, for uniform worlds are terribly boring places to live in. Boredom apart, this second interpretation of uniformity has the advantage that it allows for uniform possible worlds, or uniform possible states of the world if you prefer so.

The adaptive logic based on this idea will be called $\mathbf{IL}^r$. It suggests a completely different approach than the one followed by $\mathbf{LI}^r$. An abnormality, viz. a formula expressing a lack of uniformity, states that something has a property and something else does not. Let us at once restrict this to: an abnormality is a formula of the form $\exists(A_0 \vee \ldots \vee A_n) \wedge \exists\neg(A_0 \vee \ldots \vee A_n)$ for $A_0, \ldots, A_n \in \mathcal{A}^{f1}$ and $n \geq 0$.[16] Here is a little proof from $\{Pa\}$.

1	Pa	premise	$\emptyset$
2	$\forall xPx$	1; RC	$\{\exists Px \wedge \exists\neg Px\}$

Several simple but interesting observations may be made. From the premise Pa, $\exists xPx$ is **CL**-derivable. As $\forall xPx \vee \exists x\neg Px$ is a **CL**-theorem, either $\forall xPx$ is true or the abnormality $\exists Px \wedge \exists\neg Px$ is true. In other words, if one presupposes abnormalities to be false unless and until proven otherwise, one may derive $\forall xPx$ from Pa, unless and until some object is shown not to have the property P. Next, if Pa is a premise, $\exists Px \wedge \exists\neg Px$ is false just in case $\exists\neg Px$ is false, and this formula is **CL**-equivalent to the negation of $\forall xPx$. This provides a nice way to compare $\mathbf{IL}^r$ with $\mathbf{LI}^r$. The logic $\mathbf{LI}^r$ enables one to introduce $\forall xPx$ because $\forall xPx \vee \neg\forall xPx$ is a **CL**-theorem. $\mathbf{IL}^r$ seems more demanding. In order to introduce $\forall xPx$, one needs an instance of it, or at least $\exists xPx$. This seems a natural requirement; one does not want to introduce a generalization unless one knows that it has at least one instance.

The preceding paragraph suggests that $\mathbf{IL}^r$ is more demanding than $\mathbf{LI}^r$, but actually it is not. In the absence of an instance, a generalization is marked in a $\mathbf{LI}^r$-proof anyway—see the table in the previous section. Moreover, while $\mathbf{IL}^r$ requires an instance in order to derive a generalization, it also requires factual premises in order to derive a disjunction of abnormalities—no disjunction of $\mathbf{IL}^r$-abnormalities is a **CL**-theorem. It is also instructive to see the following. The second conjunct of $\exists Px \wedge \exists\neg Px$ is **CL**-equivalent to the negation of $\forall xPx$. While $\forall xPx$ can always be introduced on the condition $\neg\forall xPx$ in a $\mathbf{LI}^r$-proof, one moreover needs $\exists xPx$, the first conjunct of $\exists Px \wedge \exists\neg Px$, in order to introduce $\forall xPx$ on the condition $\exists Px \wedge \exists\neg Px$ in a $\mathbf{IL}^r$-proof. But precisely $\exists xPx$ is sufficient to derive the $\mathbf{IL}^r$-abnormality $\exists Px \wedge \exists\neg Px$ from the $\mathbf{LI}^r$-abnormality $\exists x\neg Px$. The same reasoning applies to more complex generalizations or to disjunctions of them.

While $\mathbf{IL}^r$ is very close to $\mathbf{LI}^r$, it is also richer. This is easily seen if the previous proof is continued as follows—$!A$ abbreviates $\exists A \wedge \exists\neg A$ to fit the proof on the page.

3	$(Pa \supset Qa) \vee (Pa \supset \neg Qa)$	1; RU	$\emptyset$
4	$\forall x(Px \supset Qx) \vee \forall x(Px \supset \neg Qx)$	3; RC	$\{!(\neg Px \vee Qx), !(\neg Px \vee \neg Qx)\}$

[16] We obtain a different formulation of the same logics if one considers every formula of the form $\exists A \wedge \exists\neg A$, with $A \in \mathcal{F}_s^{f1}$, as an abnormality. I return to this at the end of the section.

We do not know anything about the Q-hood of objects that are P. Yet, we presuppose uniformity and so presuppose that all P are Q or that all P are not Q. And indeed, $\forall x(Px \supset Qx) \vee \forall x(Px \supset \neg Qx)$ is finally derivable from $\{Pa\}$. As the premise set is normal (with respect to $\mathbf{IL}^r$), no abnormality is **CL**-derivable from it and line 4 will not be marked in any extension of the proof.[17] This is a gain over $\mathbf{LI}^r$. In a $\mathbf{LI}^r$-proof from $\{Pa\}$, the (reformulated) condition of line 4 is $\{\exists x\neg(\neg Px \vee Qx), \exists x\neg(\neg Px \vee \neg Qx)\}$ and the disjunction of the two members of the condition is **CL**-derivable from Pa, whence line 4 is marked in the $\mathbf{LI}^r$-proof.

Our new logic, $\mathbf{IL}^r$, is also richer than $\mathbf{LI}^r$ in other respects. Here is a nice example, provided to me by Mathieu Beirlaen. Reconsider $\Gamma_2 = \{Pa, \neg Pb \vee Qb\}$. In $\mathbf{LI}^r$ no generalization is finally derivable from Γ_2. Indeed, Pa entails the abnormality $\exists x\neg\neg Px$. The second premise informs one that either the abnormality $\exists x\neg Px$ or $\exists x\neg\neg Qx$ obtains. Note that both choices are on a par. The proof goes as follows—line 4 is pretty useless but is added for the sake of completeness.

1	Pa	premise	$\emptyset$	
2	$\neg Pb \vee Qb$	premise	$\emptyset$	
3	$\forall xPx$	1; RC	$\{\exists x\neg Px\}$	$\checkmark^5$
4	$\exists x\neg\neg Px$	1; RU	$\emptyset$	
5	$\exists x\neg Px \vee \exists x\neg\neg Qx$	2; RU	$\emptyset$	

So line 3 is marked in every extension of the proof from Γ because $\exists x\neg Px \in U_s(\Gamma_2)$. The situation in $\mathbf{IL}^r$ is completely different.

1	Pa	premise	$\emptyset$
2	$\neg Pb \vee Qb$	premise	$\emptyset$
3	$\forall xPx$	1; RC	$\{\exists xPx \wedge \exists x\neg Px\}$
4	$\forall xQx$	3, 2; RC	$\{\exists xPx \wedge \exists x\neg Px, \exists x\neg Qx \wedge \exists x\neg\neg Qx\}$

With respect to $\mathbf{IL}^r$, Γ_2 is normal and no *Dab*-formula is derivable from it. So both lines 3 and 4 are unmarked in all extensions of the proof.

We have thus arrived at a logic of inductive generalization, $\mathbf{IL}^r$, that has **CL** as its lower limit logic, $\{\exists(A_0 \vee \ldots \vee A_n) \wedge \exists\neg(A_0 \vee \ldots \vee A_n) \mid A_0, \ldots, A_n \in \mathcal{A}^{f1}; n \geq 0\}$ as its set of abnormalities, and Reliability as its strategy. Moreover, $\mathbf{IL}^r$ has an upper limit logic that is not the trivial logic **Tr**. Let us call this upper limit logic **UCL**, which abbreviates "uniform classical logic". It is obtained by adding to **CL** the axiom schema $\exists A \supset \forall A$.[18] The **UCL**-models are those **CL**-models in which all objects have exactly the same properties. So, where $\pi \in \mathcal{P}^r$, $v(\pi)$ is either the empty set or the rth Cartesian product of the domain, which is the set of all r-tuples of members of the domain. The variant $\mathbf{IL}^m$ is similar except that it has Minimal Abnormality as its strategy and is slightly stronger than $\mathbf{IL}^r$.

The rules of $\mathbf{IL}^r$ and $\mathbf{IL}^m$ are exactly like those of $\mathbf{LI}^r$, except that the set of abnormalities is different. $U_s(\Gamma)$ and $\Phi_s(\Gamma)$ are also defined as in the previous

[17] Reconsider the previous paragraph of the text for a moment. On the one hand no line will be marked in a proof from $\{Pa\}$. On the other hand, applying RC requires a positive instance of certain generalizations.

[18] Requiring that $A \in \mathcal{F}_s^{f1}$ is useless as is obvious from the description of the semantics that follows in the text. With or without that restriction, the axiom schema has the same effect.

chapter and the respective marking definitions for Reliability and Minimal Abnormality are literally the same. The same holds for the semantics. One defines $U_s(\Gamma)$, $\Phi_s(\Gamma)$, and $Ab(M)$, and defines the models selected by the strategy as in the previous chapter.

Let us quickly consider some aspects of the way in which the present logics may guide research. Let $\Gamma_4 = \{Pa, Qa, \neg Qb, \neg Pc\}$. Here is an $\mathbf{IL}^r$-proof from Γ_4—$!A$ abbreviates $\exists A \wedge \exists\neg A$ as before.

1	Pa	premise	$\emptyset$
2	Qa	premise	$\emptyset$
3	$\neg Qb$	premise	$\emptyset$
4	$\neg Pc$	premise	$\emptyset$
5	$\forall x(Px \supset Qx)$	2; RC	$\{!(\neg Px \vee Qx)\}$ ✓[7]
6	$(Pb \wedge \neg Qb) \vee (\neg Pb \wedge \neg Qb)$	3; RU	$\emptyset$
7	$!(\neg Px \vee Qx) \vee !(Px \vee Qx)$	2, 6; RC	$\emptyset$

As $\Gamma_4 \nvdash_{\mathbf{CL}} !(Px \vee Qx)$, line 5 is marked in all extensions of the proof. So in order to have line 5 unmarked, one needs to derive $\exists x(Px \vee Qx) \wedge \exists x \neg(Px \vee Qx)$. The first conjunct follows from 2. So we moreover need an instance of $\neg Px \wedge \neg Qx$. For this, obtaining for example the new information $\neg Pb$ or the new information $\neg Qc$ is sufficient.

Is $\mathbf{IL}^r$ or $\mathbf{IL}^m$ the optimal logic of inductive generalization? This is an odd question. Inductive generalization is a methodological matter and it would be odd that there were a single optimal and context-independent method. There are indeed alternatives, as I shall now illustrate. Actually, Γ_4 provides a nice starting point to arrive at an alternative. Presumably some would like to derive $\forall x(Px \equiv Qx)$ from it. This does not follow by $\mathbf{IL}^r$, but it follows by a slightly different adaptive logic. To even derive $\forall x(Px \supset Qx)$ at a stage and on the condition $\{!(\neg Px \vee Qx)\}$, we need an instance of $Px \supset Qx$. What is meant here is an instance in the sense in which logicians use the term, for example $Pd \supset Qd$. Note that this follows from $\neg Pd$ as well as from Qd. When philosophers of science talk about inductive generalization, they use the phrases *positive* instance and *negative* instance. By a positive instance of $\forall x(Px \supset Qx)$ they mean an instance of $Px \wedge Qx$ and by a negative instance of $\forall x(Px \supset Qx)$ they mean an instance of $Px \wedge \neg Qx$. This suggests that we consider adaptive logics that have such abnormalities as $\exists x(Px \wedge Qx) \wedge \exists x(Px \wedge \neg Qx)$—in words: there is a positive as well as a negative instance of $\forall x(Px \supset Qx)$.

One way to do this systematically is by defining the set of abnormalities as $\{\exists(A_1 \wedge \ldots \wedge A_n \wedge A_0) \wedge \exists(A_1 \wedge \ldots \wedge A_n \wedge \neg A_0) \mid A_0, A_1, \ldots, A_n \in \mathcal{A}^{f1}; n \geq 0\}$.[19] As the abnormalities are long, I shall abbreviate $\exists(A_1 \wedge \ldots \wedge A_n \wedge A_0) \wedge \exists(A_1 \wedge \ldots \wedge A_n \wedge \neg A_0)$ in proofs as $A_1 \wedge \ldots \wedge A_n \wedge \pm A_0$.

This approach is sufficiently general because every generalization, as restricted in Section 3.1, is **CL**-equivalent to a conjunction of formulas of the form $\forall((A_1 \wedge \ldots \wedge A_n) \supset A_0)$ (where the metavariables denote members of $\mathcal{A}^{f1}$). So a generalization G will be derivable just in case the formulas are derivable that have the specified form and are **CL**-entailed by G. This is precisely as we want it.

[19]If $n = 0$, the formula reduces to $\exists A_0 \wedge \exists\neg A_0$. As before, I do not rule out that $(n + 1)$ different variables occur in the abnormalities, but I restrict the attention in the rest of the chapter to generalizations containing one variable only.

Note that the three formulas in the left column of the following table have the required form and are **CL**-equivalent.

$\forall x((Px \wedge Qx) \supset Rx)$	$\exists x(Px \wedge Qx \wedge Rx) \wedge \exists x(Px \wedge Qx \wedge \neg Rx)$
$\forall x((Px \wedge \neg Rx) \supset \neg Qx)$	$\exists x(Px \wedge \neg Rx \wedge \neg Qx) \wedge \exists x(Px \wedge \neg Rx \wedge \neg\neg Qx)$
$\forall x((\neg Rx \wedge Qx) \supset \neg Px)$	$\exists x(\neg Rx \wedge Qx \wedge \neg Px) \wedge \exists x(\neg Rx \wedge Qx \wedge \neg\neg Px)$

In the right column, I list the corresponding abnormalities. The second conjunct of the abnormalities is **CL**-equivalent to the negation of the corresponding generalization (and all these second conjuncts are obviously **CL**-equivalent to each other). As was already mentioned in the previous chapter, a formula A is derivable on a condition Δ in a proof from Γ just in case $A \vee Dab(\Delta)$ is derivable on the empty condition in that proof. So in order to derive the generalization by the rule RC, the premises need to provide the information contained in the first conjunct of the abnormality. As the three generalizations are **CL**-equivalent to each other, they may be derived from each other by the rule RU. So the generalizations can be derived on three different conditions.

The following table shows that the matter is actually very simple.

$\forall x(\neg Px \vee \neg Qx \vee \underline{Rx})$	$\exists x(Px \wedge Qx \wedge Rx) \wedge \exists x(Px \wedge Qx \wedge \neg Rx)$
$\forall x(\neg Px \vee \underline{\neg Qx} \vee Rx)$	$\exists x(Px \wedge \neg Rx \wedge \neg Qx) \wedge \exists x(Px \wedge \neg Rx \wedge \neg\neg Qx)$
$\forall x(\underline{\neg Px} \vee \neg Qx \vee Rx)$	$\exists x(\neg Rx \wedge Qx \wedge \neg Px) \wedge \exists x(\neg Rx \wedge Qx \wedge \neg\neg Px)$

The three generalizations from the previous table are all **CL**-equivalent to the (three times repeated) formula in the left column of this table. Select a disjunct (the underlined one) of this formula. The first conjunct of the abnormality is the existential closure of the conjunction of the selected disjunct together with the negation of the other disjuncts. The second conjunct of the abnormality is the existential closure of the conjunction of the negation of all disjuncts. The number of disjuncts obviously determines the number of different abnormalities.

Incidentally, the definition of the abnormalities does not require that the members of $\mathcal{A}^{f1}$ are different, but actually there is no need to require so. If, for example, one replaces Q by P in the tables, one abnormality is equivalent to the abnormality corresponding to $\forall x(Px \supset Rx)$. The other two abnormalities are **CL**-contradictions, whence no consistent set of singular premises enables one to derive $\forall x((Px \wedge Px) \supset Rx)$ on those conditions.[20]

So I have described and clarified two further logics. Let us call them $\mathbf{G}^r$ and $\mathbf{G}^m$. Their lower limit is obviously **CL**, their set of abnormalities is the one defined five paragraphs ago, and their strategies are respectively Reliability and Minimal Abnormality.

Let us now return to the premise set Γ_4 and check that the new logics assign the consequence $\forall x(Px \equiv Qx)$ to it.

1	Pa	premise	$\emptyset$
2	Qa	premise	$\emptyset$
3	$\neg Qb$	premise	$\emptyset$
4	$\neg Pc$	premise	$\emptyset$
5	$\forall x(Px \supset Qx)$	1, 2; RC	$\{Px \wedge \pm Qx\}$

[20]Other replacements lead to different results; if the generalization is a **CL**-theorem, the abnormalities are all **CL**-contradictions. There is nothing interesting about this and if one considers all replacements, including the replacement of R by $\neg P$ or $\neg Q$, the whole matter turns out boringly symmetric.

6	$\forall x(Qx \supset Px)$	1, 2; RC	$\{Qx \wedge \pm Px\}$
7	$\forall x(Qx \equiv Px)$	5, 6; RU	$\{Px \wedge \pm Qx, Qx \wedge \pm Px\}$
8	$\exists x Px \wedge \exists x \neg Px$	1, 4; RU	$\emptyset$
9	$\exists x Qx \wedge \exists x \neg Qx$	2, 3; RU	$\emptyset$

The only minimal *Dab*-consequences of Γ_4 are 8 and 9. So lines 5–7 will be unmarked in all extensions of this proof. Obviously the proof is also a correct $\mathbf{G}^m$-proof and no line is marked in it or will be marked in any extension of it.

The upper limit logic of $\mathbf{G}^r$ and $\mathbf{G}^m$ is still **UCL**. This is easily seen if one realizes that a model verifies an abnormality as soon as two elements of the domain do not share a primitive property. Premise set Γ_4, for example, is abnormal because $\exists x Px \wedge \exists x \neg Px$ and $\exists x Qx \wedge \exists x \neg Qx$ are *Dab*-consequences of it.

It can be shown that the logics **IL**-family at least as strong than the corresponding logics of the **LI**-family and in general stronger. Thus $Cn_{\mathbf{LI}^r}(\Gamma) \subseteq Cn_{\mathbf{IL}^r}(\Gamma)$ holds for all Γ and the inclusion is proper for some Γ. The same holds for the relation between $Cn_{\mathbf{IL}^m}(\Gamma)$ and $Cn_{\mathbf{G}^m}(\Gamma)$, but not for the relation between $Cn_{\mathbf{LI}^r}(\Gamma)$ and $Cn_{\mathbf{G}^r}(\Gamma)$. An example is provided by the premise set $\Gamma_5 = \{Pa, Qb, Rb, Qc, \neg Rc\}$. While $\forall x(Px \vee Qx) \in Cn_{\mathbf{LI}^r}(\Gamma_5)$ and $\forall x(Px \vee Qx) \in Cn_{\mathbf{IL}^r}(\Gamma_5)$, $\forall x(Px \vee Qx) \notin Cn_{\mathbf{G}^r}(\Gamma_5)$. The reason lies in the absence, in Γ_5, of a positive instance of $\forall x(\neg Px \supset Qx)$ as well as of $\forall x(\neg Qx \supset Px)$.

Some pages ago, I claimed that removing the restriction on the abnormalities of $\mathbf{IL}^r$ and $\mathbf{IL}^m$ leads to a different formulation of the same logics. This is not an important claim, so let me just illustrate it for $\mathbf{IL}^r$. In the present formulation, one can introduce the generalization $\forall x((Px \vee Qx) \supset Rx)$ on the condition $\{\exists x(\neg Px \vee Rx) \wedge \exists x \neg(\neg Px \vee Rx), \exists x(\neg Qx \vee Rx) \wedge \exists x \neg(\neg Qx \vee Rx)\}$. What is needed to do so is either an instance of Rx or an instance of $\neg Px$ as well as an instance of $\neg Qx$. If the restriction is removed, one can *also* introduce the generalization on the condition $\{\exists x((\neg Px \wedge \neg Qx) \vee Rx) \wedge \exists x \neg((\neg Px \wedge \neg Qx) \vee Rx)\}$. What is needed to do so is either an instance of Rx or an instance of $\neg Px \wedge \neg Qx$—so the *same* object should now be non-P as well as non-Q. The crucial question is whether it is possible that $\exists x(\neg Px \vee Rx) \wedge \exists x \neg(\neg Px \vee Rx)$ or $\exists x(\neg Qx \vee Rx) \wedge \exists x \neg(\neg Qx \vee Rx)$ is a disjunct of a *Dab*-consequence of the premises whereas $\exists x((\neg Px \wedge \neg Qx) \vee Rx) \wedge \exists x \neg((\neg Px \wedge \neg Qx) \vee Rx)$ is not. Suppose that, for example, $\exists x(\neg Px \vee Rx) \wedge \exists x \neg(\neg Px \vee Rx)$ is a **CL**-consequence of the premises. So the premises contain either (i) an instance of Rx and an instance of $Px \wedge \neg Rx$ or (ii) an instance of $\neg Px$ and an instance of $Px \wedge \neg Rx$. In case (i), $\exists x((\neg Px \wedge \neg Qx) \vee Rx) \wedge \exists x \neg((\neg Px \wedge \neg Qx) \vee Rx)$ is a **CL**-consequence of the premises. In case (ii), it is also a **CL**-consequence of the premises unless there is no instance of $\neg Px \wedge \neg Qx$. In that case, however, the generalization cannot be introduced on the condition $\exists x((\neg Px \wedge \neg Qx) \vee Rx) \wedge \exists x \neg((\neg Px \wedge \neg Qx) \vee Rx)$, as we saw before. This is the general situation. If, by removing the restriction, a generalization can be introduced on a further condition, and its present condition causes the line to be marked, then the further condition will have the same effect.

It is useful to illustrate that $\mathbf{G}^r$-proofs from some premise sets may proceed in a rather unexpected way. Consider $\Gamma_6 = \{Pa, \neg Pb, Qb\}$. The disjunction $\forall x Qx \vee \forall x(Px \supset \neg Qx)$ is a final $\mathbf{G}^r$-consequence of this premise set, but neither disjunct is. This causes the proof to require some ingenuity. Let us begin by deriving all minimal *Dab*-formulas and next take some preparatory steps.

1	Pa	Prem	$\emptyset$
2	$\neg Pb$	Prem	$\emptyset$
3	Qb	Prem	$\emptyset$
4	$\pm Px$	1, 2; RU	$\emptyset$
5	$\pm Qx \vee (Qx \wedge \pm Px)$	1, 2, 3; RU	$\emptyset$
6	$\exists x(\neg Px \wedge Qx)$	2, 3; RU	$\emptyset$
7	$\forall x Qx \vee \exists x \neg Qx$	RU	$\emptyset$
8	$\forall x Qx \vee \exists x(Px \wedge \neg Qx) \vee \exists x(\neg Px \wedge \neg Qx)$	7; RU	$\emptyset$

Note that 7 is a **CL**-theorem and that its second disjunct is the first conjunct of an abnormality. By splitting it up at line 8 we obtain a disjunction the second and third disjunct of which are sides of abnormalities. In the sequel of the proof, below, both disjuncts are combined with other formulas into an abnormality. The first other formula is 6, which comes from the premises. The second other formula is the second disjunct of another **CL**-theorem, viz. $\forall x(Px \supset \neg Qx) \vee \exists x(Px \wedge Qx)$.

9	$\forall x Qx \vee \exists x(Px \wedge \neg Qx)$	6, 8; RC	$\{\neg Px \wedge \pm Qx\}$
10	$\forall x(Px \supset \neg Qx) \vee \exists x(Px \wedge Qx)$	RU	$\emptyset$
11	$\forall x Qx \vee \forall x(Px \supset \neg Qx)$	9, 10; RC	$\{\neg Px \wedge \pm Qx, Px \wedge \pm Qx\}$

The transition from 9 and 10 to 11 is an application of the $A \vee B, C \vee D \vdash_{\mathbf{CL}} A \vee C \vee (B \wedge D)$. The formula corresponding to $B \wedge D$ is an abnormality and so is pushed to the condition. As 4 and 5 are the only minimal *Dab*-consequences of Γ_6, line 11 will not be marked in any extension of the proof.

The reader should not be misled by this example. The premise set displays nearly no variety. Incidentally, $\forall x(Px \vee Qx)$ is also a final $\mathbf{G}^r$-consequence of it. The final consequences tell us less about the world than about the presumptions of our methods: our willingness to minimize inductive abnormalities. An even more striking effect of those presumptions is that $\forall x Qx \vee \forall x \neg Qx$ is a final $\mathbf{G}^r$-consequence of $\{Pa\}$.

The reader might think that $\forall x Qx \vee \forall x \neg Qx$ is also a final $\mathbf{G}^r$-consequence of $\{Pa, \neg Pb\}$. This is mistaken because these premises **CL**-entail the *Dab*-formula $\pm Qx \vee (Qx \wedge \pm Px) \vee (\neg Qx \wedge \pm Px)$. This brings us right to the topic of the following section.

3.5 Combined Adaptive Logics

Do not think that you have seen all adaptive logics of inductive generalization. The fun only starts. In this section, I introduce *combined* adaptive logics, logics that are specific combinations of fragments of the adaptive logics we have met so far. Uncombined adaptive logics will from now on be called *simple* adaptive logics, and when I write "adaptive logic" without adding either "simple" or "combined", I mean a simple adaptive logic, unless where the context indicates the contrary.

Let us begin again with an extremely unsophisticated example, a $\mathbf{G}^r$-proof from $\Gamma_7 = \{Pa, Qa, \neg Qb\}$.

1	Pa	premise	$\emptyset$
2	Qa	premise	$\emptyset$

3	$\neg Qb$	premise	$\emptyset$	
4	$\forall x Qx$	2; RC	$\{\pm Qx\}$	$\checkmark^5$
5	$\pm Qx$	2, 3; RU	$\emptyset$	
6	$\forall x Px$	1; RC	$\{\pm Px\}$	$\checkmark^7$
7	$\pm Px \vee (Px \wedge \pm Qx)$	1, 2, 3; RU	$\emptyset$	
8	$\forall x(Qx \supset Px)$	1, 2; RC	$\{Qx \wedge \pm Px\}$	

The last line will not be marked in any extension of the proof. If you do not see at once that 7 follows, note that either $\neg Pb$ or Pb. In the first case, $\exists x Px \wedge \exists x \neg Px$ holds true; in the second case, $\exists x(Px \wedge Qx) \wedge \exists x(Px \wedge \neg Qx)$ holds true.[21]

Some may argue that $\forall x Px$ should be derivable from the premises. If P and $\neg P$ have the same logical width,[22] the generalization $\forall x Px$ is more general than $\forall x(Px \supset Qx)$ and Popper [Pop35] has argued that more general hypotheses should be given precedence—for being tested, as Popper sees it. Given the data, $\forall x Px$ has more potential falsifiers, viz. every instance of $\neg Px$, than $\forall x(Px \supset Qx)$, which is only falsified by instances of $Px \wedge \neg Qx$.

So let us modify $\mathbf{G}^r$ in such a way that $\forall x Px$ is a consequence of Γ_7. I first perform the technical trick and next explain it. We begin by splitting up the set of abnormalities, Ω, in subsets Ω^i which we define as $\Omega^i = \{\exists x(A_1 \wedge \ldots \wedge A_i \wedge A_0) \wedge \exists x(A_1 \wedge \ldots \wedge A_i \wedge \neg A_0) \mid A_0, A_1, \ldots, A_i \in \mathcal{A}^{f1}\}$. So $\exists x Px \wedge \exists x \neg Px \in \Omega^0$, $\exists x(Px \wedge Qx) \wedge \exists x(Px \wedge \neg Qx) \in \Omega^1$, etc. Let us call i the *degree* of the abnormalities comprised in Ω^i.[23] Next we define, for every $i \in \mathbb{N}$, $\Omega^{(i)} = \Omega^0 \cup \ldots \cup \Omega^i$. Finally we define, for every $i \in \mathbb{N}$, the adaptive logic $\mathbf{G}^r_{(i)}$ in the same way as $\mathbf{G}^r$ was defined,[24] except that the set of abnormalities of $\mathbf{G}^r_{(i)}$ is not Ω but $\Omega^{(i)}$. So $\mathbf{G}^r_{(0)}$ is just like $\mathbf{G}^r$, except that its set of abnormalities is $\Omega^{(0)}$ instead of Ω. Similarly for $\mathbf{G}^r_{(1)}$, except that its set of abnormalities is $\Omega^{(1)}$, and so on.

Let me pause a moment here. What is the effect, for example, on Γ_7? Consider again the little proof. Line 6 was marked because its condition overlaps with the disjuncts of the *Dab*-formula derived at line 7. What becomes of this proof if we transform it to a $\mathbf{G}^r_{(0)}$-proof from Γ_7? First of all, line 8 cannot be derived any more because its condition comprises a formula that is not a member of $\Omega^{(0)}$. So suppose we delete line 8. Line 7 is obviously still derivable—it is a **CL**-consequence of the premises—but it is not a *Dab*-formula. Indeed, its second disjunct has degree 1 and so is not a member of $\Omega^{(0)}$. So line 6 is not marked in the present $\mathbf{G}^r_{(0)}$-proof. Moreover, it will not be marked in any extension of the proof. The only minimal *Dab*-formula that is derivable from Γ_7 is $\pm Qx$. As $\pm Qx$ is derived in the proof, line 6 will be unmarked in all extensions of the

[21] The generalization $\forall x(Qx \supset Px)$ is also finally derivable from Γ_7 by $\mathbf{LI}^r$ and $\mathbf{IL}^r$. By $\mathbf{G}^r$ the disjunction $\forall x(\neg Px \supset Qx) \vee \forall x(Px \supset Qx)$ is moreover finally derivable. For the variants that have Minimal Abnormality as their strategy, $\forall x Px \vee \forall x(Px \supset Qx)$ is also finally derivable.

[22] This means that, as far as *logical* considerations are concerned, it is equally likely that an object is P than that it is $\neg P$. For example, it takes a large number of properties for an object to be human, whereas the absence of any of these properties is sufficient to be non-human; so the logical width of "human" is much smaller than the logical width of "non-human".

[23] I shall use the term degree in connection with several combined adaptive logics. The meaning of the term will be contextual, but the degree of an abnormality will always have a similar function.

[24] I apologize for the fact that the superscript of Ω suddenly changes into the subscript of the logic. The superscript of $\Omega^{(i)}$ will correspond to the superscripts in $U^{(i)}_s(\Gamma)$ and in $\Phi^{(i)}_s(\Gamma)$.

proof. So $\forall xPx$ is finally $\mathbf{G}^r_{(0)}$-derivable from Γ_7.

Of course, we are not home yet. The logic $\mathbf{G}^r_{(0)}$ will not enable one to derive any generalization of the form $\forall((A_1 \wedge \ldots \wedge A_n) \supset A_0)$ unless when $\forall A_0$ is derivable or $\forall\neg A_i$ is derivable for $0 \geq i \geq n$. So in order to solve the general problem, we cast the logics $\mathbf{G}^r_{(i)}$ into a single entity, which I shall call $\mathbf{HG}^r$—the first letter refers to the fact that the logic is a combination of a hierarchy of logics. The $\mathbf{HG}^r$-consequence set of Γ is defined as follows:

$$Cn_{\mathbf{HG}^r}(\Gamma) = Cn_{\mathbf{CL}}(Cn_{\mathbf{G}^r_{(0)}}(\Gamma) \cup Cn_{\mathbf{G}^r_{(1)}}(\Gamma) \cup \ldots). \tag{3.2}$$

Note that the union of consequence sets is closed under the lower limit logic **CL**. I shall have to return to this when we come to the proofs.

Some readers might get scared by (3.2). For one thing, $\mathbf{HG}^r$ is defined in terms of an infinity of logics. But this is not a problem. It is not a theoretical problem because all those logics are well-defined and so is $\mathbf{HG}^r$. It is not a practical problem either because the premise sets to which we want to apply $\mathbf{HG}^r$ is forcibly a finite set of singular data and people applying $\mathbf{HG}^r$ will only be interested in hypotheses built from predicates that occur in the data. So every application will require that at most a finite number of $\mathbf{G}^r_{(i)}$ logics are invoked.[25]

But there is another problem. What is a $\mathbf{HG}^r$-proof supposed to look like? Suppose you want to derive a prediction from the data together with inductively inferred generalizations. Is it possible to do so in a single proof? The generalizations may require that different $\mathbf{G}^r_{(i)}$ logics are invoked and all of these have different rules and different marking definitions—remember that the rules and the marking definition refer to the set of abnormalities. Well, that's not a problem either. Let me explain.

The point is of great importance. The adaptive logic $\mathbf{HG}^r$ is a combined one. It uses a specific and very simple kind of combination of infinitely many adaptive logics. For this logic, as for other combined adaptive logics,[26] the proofs are in principle not more complex than for simple adaptive logics. Moreover, the way to approach proofs of combined adaptive logics is always the same. First, one allows the application of the *rules* of *all* combined logics in the combined proof. Next, one adjusts the marking definition in such a way that it leads to a correct combined result.

To remove any possible confusion, let me spell out the matter. Let a $Dab^{(i)}$-formula be a disjunction of members of $\Omega^{(i)}$. The rules for $\mathbf{G}^r_{(i)}$ (as well as for $\mathbf{G}^m_{(i)}$) are the following.

[25]In Section 3.6 background knowledge will be considered. Even there, the set of predicates will be finite.

[26]To be precise, this holds for all combined adaptive logics that were studied so far.

Prem	If $A \in \Gamma$:	$\ldots$	$\ldots$
		A	$\emptyset$
RU	If $A_1, \ldots, A_n \vdash_{\mathbf{CL}} B$:	A_1	Δ_1
		$\ldots$	$\ldots$
		A_n	Δ_n
		B	$\Delta_1 \cup \ldots \cup \Delta_n$
RC$^{(i)}$	If $A_1, \ldots, A_n \vdash_{\mathbf{CL}} B \vee Dab^{(i)}(\Theta)$:	A_1	Δ_1
		$\ldots$	$\ldots$
		A_n	Δ_n
		B	$\Delta_1 \cup \ldots \cup \Delta_n \cup \Theta$

The only reference to i occurs in RC$^{(i)}$, viz. in the metalinguistic formula $B \vee Dab^{(i)}(\Theta)$. The rules Prem and RU are identical for all logics $\mathbf{G}^r_{(i)}$.

Proofs of combined logics have a very specific characteristic. Conditions of lines may contain abnormalities from any $\Omega^{(i)}$, and actually from several $\Omega^{(i)}$. In applying RU and RC$^{(i)}$ to lines with such conditions the full condition is carried over.[27] Needless to say, this will have to be taken into account when we phrase the marking definition.

Another problem has to be solved: the consequence set has to be closed under **CL** as is clear from (3.2). Let us postpone this complication for a moment and turn the $\mathbf{G}^r$-proof from Γ_7 into a $\mathbf{HG}^r$-proof.

1	Pa	premise	$\emptyset$	
2	Qa	premise	$\emptyset$	
3	$\neg Qb$	premise	$\emptyset$	
4	$\forall x Qx$	2; RC$^{(0)}$	$\{\pm Qx\}$	$\checkmark^5$
5	$\pm Qx$	2, 3; RU	$\emptyset$	
6	$\forall x Px$	1; RC$^{(0)}$	$\{\pm Px\}$	
7	$\pm Px \vee (Px \wedge \pm Qx)$	1, 2, 3; RU	$\emptyset$	
8	$\forall x(Qx \supset Px)$	1, 2; RC$^{(1)}$	$\{Qx \wedge \pm Px\}$	

This simple proof contains only applications of two combining logics, $\mathbf{G}^r_{(0)}$ and $\mathbf{G}^r_{(1)}$. Remember that the superscript of RC refers to them. Lines 4 and 6 are written by application of the conditional rule of $\mathbf{G}^r_{(0)}$. Given the specific combination we are considering, see (3.2), lines 4 and 6 can be written by application of the conditional rule of any $\mathbf{G}^r_{(i)}$ for $i \geq 0$—I choose the lowest i because that corresponds to the degree of the condition of the line, as becomes clear below. As it stands, line 8 cannot be written by an application of the conditional rule of $\mathbf{G}^r_{(0)}$, but is justifiable by the conditional rule of any $\mathbf{G}^r_{(i)}$ for $i \geq 1$, so I choose $\mathbf{G}^r_{(1)}$. Of course, as line 6 is unmarked, line 8 could be derived from that by RU on the condition of line 6.

I now set out to the specify the marking definition, still neglecting the closure under **CL**. Let the *degree of a condition* be the maximal degree of its members. Incidentally, it is very easy to decide which is the lowest i for which $\mathbf{G}^r_{(i)}$ enables one to obtain a line: i is the degree of the condition of the line. Remember that $Cn_{\mathbf{HG}^r}(\Gamma)$ is defined, in (3.2), as the closure of a union of consequence

[27] The example proofs that follow do not sufficiently illustrate this feature, but the proof from Γ_{10} does (see page 236).

sets. So if a formula is finally derivable by one of the $\mathbf{G}^r_{(i)}$ logics, it is finally $\mathbf{HG}^r$-derivable. Moreover, in the present context every $Dab^{(i)}$-formula is also a $Dab^{(i+1)}$-formula. So if a line is marked in view of $\mathbf{G}^r_{(i)}$, it is also marked in view of $\mathbf{G}^r_{(i+1)}$ but not necessarily in view of $\mathbf{G}^r_{(i-1)}$ (if $i \geq 1$). So if i is the maximal degree of a member of the condition of the line and the line is unmarked in view of $\mathbf{G}^r_{(i)}$, then the line should be unmarked in the $\mathbf{HG}^r$-proof. Finally, where $Dab^{(n)}(\Delta_1), \ldots, Dab^{(n)}(\Delta_m)$ are the minimal $Dab^{(n)}$-formulas that are derived at stage s of a $\mathbf{HG}^r$-proof from Γ, $U^{(n)}_s(\Gamma) =_{df} \Delta_1 \cup \ldots \cup \Delta_m$. The *provisional* (or rather partial) marking definition reads as follows.

Definition 3.5.1 *Line l is marked at stage s iff, where Δ is its condition and d is the degree of Δ, $\Delta \cap U^{(d)}_s(\Gamma) \neq \emptyset$.*

So line 6 of the $\mathbf{HG}^r$-proof is unmarked because its condition has degree 0, $U^{(0)}_8(\Gamma_7) = \{\pm Qx\}$, and $\{\pm Px\} \cap \{\pm Qx\} = \emptyset$. Note that $\pm Px \in U^{(1)}_8(\Gamma_7)$ but that line 6 is nevertheless unmarked in view of Definition 3.5.1 because the degree of the condition of line 6 is 0.

In which way should the closure by **CL** be built into the proofs? Several roads are open, but the idea is always the same: whatever can be derived by **CL** from unmarked lines, is itself unmarked. So let us introduce the following rule.

RU*	If $A_1, \ldots, A_n \vdash_{\mathbf{CL}} B$:	A_1	Δ_1
		$\ldots$	$\ldots$
		A_n	Δ_n
		B	$*$

The $*$ in the condition of the local conclusion may be replaced by a set of sets of conditions, viz. $\{\Delta_1, \ldots, \Delta_n\}$. However, there is no need to specify these sets. The marking definition, spelled out below, simply states that a line which has $*$ as its condition is marked iff any of the lines mentioned in the justification of the line are marked.

Definition 3.5.2 *Line l is marked at stage s iff, (i) where Δ is its condition and d is the degree of Δ, $\Delta \cap U^{(d)}_s(\Gamma) \neq \emptyset$ and (ii) where $*$ is its condition, a line mentioned in the justification of l is marked at stage s. (Marking for* $\mathbf{HG}^r$*.)*

Let us consider a simple example of a proof in which the rule RU* is applied and (ii) of the marking definition has effect. The premise set is $\Gamma_8 = \{Pa \wedge Ra, \neg Rb \wedge \neg Pb, Qc \wedge Pd\}$.

1	$Pa \wedge Ra$	premise	$\emptyset$	
2	$\neg Rb \wedge \neg Pb$	premise	$\emptyset$	
3	$Qc \wedge Pd$	premise	$\emptyset$	
4	$\forall x Px$	3; RC$^{(0)}$	$\{\pm Px\}$	✓[9]
5	$\forall x Qx$	3; RC$^{(0)}$	$\{\pm Qx\}$	
6	$\forall x(Px \supset Rx)$	1; RC$^{(1)}$	$\{Px \wedge \pm Rx\}$	
7	$Pc \wedge (Qc \wedge Rc)$	3, 4, 6; RU*	$*$	✓[9]
8	$Pd \wedge (Qd \wedge Rd)$	3, 5, 6; RU*	$*$	
9	$\pm Px$	1, 2; RU	$\emptyset$	

Line 4 is marked in view of line 9. Line 7 is marked as soon as line 4 is marked because 4 occurs in the justification of line 7. Line 8 is unmarked because all lines mentioned in its justification are unmarked. Actually, 5 and 8 are final consequences of Γ_8. Another, presumably unexpected, application of RU* is presented on page 236.

The logic $\mathbf{HG}^m$ is obtained in a similar way. One defines $\Phi_s^{(d)}(\Gamma)$ and adjusts the marking definition.[28] Moreover, the same combination may be realized for the logics of the **LI**-family and for those of the **IL**-family. Remember, in this connection, that the abnormalities of the **LI**-family were redefined at the beginning of Section 3.4. The simplified abnormalities have the form $\exists\neg(A_0 \vee \ldots \vee A_n)$ with $A_0, \ldots, A_n \in \mathcal{A}^{f1}$. Call n the degree of these abnormalities. Where the set of these abnormalities is Ω, consider the subsets Ω^i that comprise the members of Ω that have degree i. Define, for every $i \in \mathbb{N}$, $\Omega^{(i)} = \Omega^0 \cup \ldots \cup \Omega^i$. The adaptive logic $\mathbf{LI}^r_{(i)}$ is exactly like $\mathbf{LI}^r$ except that the set of abnormalities of $\mathbf{LI}^r_{(i)}$ is not Ω but $\Omega^{(i)}$. Finally, we define the $\mathbf{HLI}^r$ consequence set.

$$Cn_{\mathbf{HLI}^r}(\Gamma) = Cn_{\mathbf{CL}}(Cn_{\mathbf{LI}^r_{(0)}}(\Gamma) \cup Cn_{\mathbf{LI}^r_{(1)}}(\Gamma) \cup \ldots) \tag{3.3}$$

The way in which $\mathbf{HLI}^r$-proofs are obtained is wholly similar to the way in which $\mathbf{HG}^r$-proofs were obtained. We define $Dab^{(i)}$-formulas and $U_s^{(i)}(\Gamma)$ (for all i). Again, one applies the rules of all combining logics and defines marking as in Definition 3.5.2. I leave it as an easy exercise for the reader to define $\mathbf{HLI}^m$. A further easy exercise is to define $\mathbf{HIL}^r$ and $\mathbf{HIL}^m$. So now we have six simple logics of inductive generalization and six combined ones. All the combined ones assign $\forall x Px$ as a consequence to Γ_7, which is as desired. The reason for this is always the same. The condition on which $\forall x Px$ is derived has degree 0,[29] whereas the sole member of the condition is not a disjunct of any $Dab^{(0)}$-consequence of Γ_7. It is a disjunct of a $Dab^{(1)}$-consequence of Γ_7, but that has no effect on conditions of degree 0. So the line at which $\forall x Px$ is derived is unmarked as soon as all minimal $Dab^{(0)}$-consequences of Γ_7 have been derived in the proof—there is only one $Dab^{(0)}$-consequence of Γ_7, for example $\pm Qx$ in the $\mathbf{HG}^r$-proof.

The underlying idea is still minimal joint incompatibility. What is special for the combined adaptive logics is that this idea is not applied to all generalizations together, but to sets of them. First one applies the idea to the most informative generalizations, those corresponding to abnormalities of degree zero. Next to the most informative and the second most informative generalizations together, and so on. In this way, more informative generalizations are derivable than in the case of the simple logics.

While the logics of the **H**-group form a nice enrichment with respect to their simple counterparts, a little reflection shows that one might combine the separate logics $\mathbf{LI}^r_{(i)}$, $\mathbf{IL}^r_{(i)}$, and $\mathbf{G}^r_{(i)}$ in a more profitable way. For example, for the **LI**-family the combination would look as follows.

$$Cn_{\mathbf{CLI}^r}(\Gamma) = \ldots(Cn_{\mathbf{LI}^r_{(2)}}(Cn_{\mathbf{LI}^r_{(1)}}(Cn_{\mathbf{LI}^r_{(0)}}(\Gamma))))\ldots \tag{3.4}$$

The last "..." represents only right parentheses. The logic $\mathbf{CLI}^m$ is defined in a similar way from the $\mathbf{LI}^m_{(i)}$ logics. The logics $\mathbf{CIL}^r$, $\mathbf{CIL}^m$, $\mathbf{CG}^r$, and $\mathbf{CG}^m$ are

[28] I skip this marking definition for now. It is spelled out in Section 6.2.4.

[29] The specific abnormality that occurs in the condition is different for the three logics.

also defined in a similar way, varying the combining logics and their strategy.

As the logics of the **H**-group, those of the **C**-group consider different layers of abnormalities and the generalizations derivable at those layers—the layers obviously refer to the maximal degree of abnormalities considered. There is a big difference, however, between both families. The logics of the **H**-group consider all the layers with respect to the original premise set Γ. The logics of the **C**-group 'add' the generalizations of a layer to the premises before moving on to the next layer. This is obvious from Definition 3.4. $\mathbf{LI}^r_{(1)}$ is not applied to Γ but to $Cn_{\mathbf{IL}^r_{(0)}}(\Gamma)$, $\mathbf{LI}^r_{(2)}$ to $Cn_{\mathbf{LI}^r_{(1)}}(Cn_{\mathbf{LI}^r_{(0)}}(\Gamma))$, and so on. In this way one obtains different abnormalities at the higher layers; certain *Dab*-formulas will have less disjuncts or will simply not be derivable any more. I shall now illustrate this for the **IL**-family. The example will also clarify the way in which the minimal *Dab*-formulas of the different layers are defined.

Consider the premise set $\Gamma_9 = \{Pa, Ra, \neg Pb, \neg Rb, Qc\}$ which nicely illustrates the difference between $\mathbf{HIL}^r$ and $\mathbf{CIL}^r$. Consider first the $\mathbf{HIL}^r$-proof.

1	Pa	premise	$\emptyset$	
2	Ra	premise	$\emptyset$	
3	$\neg Pb$	premise	$\emptyset$	
4	$\neg Rb$	premise	$\emptyset$	
5	Qc	premise	$\emptyset$	
6	$\forall x Qx$	5; RC$^{(0)}$	$\{!Qx\}$	
7	$\forall x(Rx \supset Px)$	1, 2; RC$^{(1)}$	$\{!(\neg Rx \vee Px)\}$	✓[11]
8	$\neg Pc \vee Pc$	RU	$\emptyset$	
9	$\neg Rc \vee Rc$	RU	$\emptyset$	
10	$(\neg Rc \wedge Qc) \vee (Rc \wedge \neg Pc) \vee (Pc \wedge Qc)$			
		5, 8, 9; RU	$\emptyset$	
11	$!(Rx \vee \neg Qx) \vee !(\neg Rx \vee Px) \vee !(\neg Px \vee \neg Qx)$			
		1, 2, 3, 10; RU	$\emptyset$	

Line 6 is unmarked in all extensions of the proof because the only $Dab^{(0)}$-formulas derivable from Γ_9 are $!Px$ and $!Rx$. Line 7, however, has a condition of degree 1 and $!(\neg Rx \vee Px)$ is a disjunct of several $Dab^{(1)}$-formulas, one of which is derived at line 11. So line 7 is marked. Actually, $\forall x(Rx \supset Px)$ is neither a final $\mathbf{HIL}^r$-consequence nor a final $\mathbf{HIL}^m$-consequence of Γ_9.

We now move to $\mathbf{CIL}^r$. Lines 1–5 are identical to those of the previous proof. I repeat the other lines from the previous proof, and next extend it. The extension proceeds slowly for the sake of clarity.

invoegen als op achterkant van blad

6	$\forall x Qx$	5; RC$^{(0)}$	$\{!Qx\}$
7	$\forall x(Rx \supset Px)$	1, 2; RC$^{(1)}$	$\{!(\neg Rx \vee Px)\}$
8	$\neg Pc \vee Pc$	RU	$\emptyset$
9	$\neg Rc \vee Rc$	RU	$\emptyset$
10	$(\neg Rc \wedge Qc) \vee (Rc \wedge \neg Pc) \vee (Pc \wedge Qc)$		
		5, 8, 9; RU	$\emptyset$
11	$!(Rx \vee \neg Qx) \vee !(\neg Rx \vee Px) \vee !(\neg Px \vee \neg Qx)$		
		1, 2, 3, 10; RU	$\emptyset$
12	Qa	6; RU	$\{!Qx\}$
13	$Pa \wedge Qa$	1, 12; RU	$\{!Qx\}$
14	$!(\neg Px \vee \neg Qx)$	3, 13; RU	$\{!Qx\}$

Again the $Dab^{(0)}$-formulas derivable from Γ_9 are $!Px$ and $!Rx$, whence line 6 is unmarked. At stage 14 of the proof, $!(Rx \vee \neg Qx) \vee !(\neg Rx \vee Px) \vee !(\neg Px \vee \neg Qx)$ is not a minimal *Dab*-formula any more. Indeed, $Dab^{(1)}$-formulas are members of $Cn_{\mathbf{IL}^r_{(0)}}(\Gamma)$, which is closed under the lower limit logic **CL**. So line 14, which is unmarked, establishes that $!(\neg Px \vee \neg Qx)$ is a $Dab^{(1)}$-consequence of Γ_9. Incidentally, $!(Rx \vee \neg Qx)$ is also a $Dab^{(1)}$-consequence of Γ_9, whereas $!(\neg Rx \vee Px) \notin U^{(1)}(Cn_{\mathbf{IL}^r_{(0)}}(\Gamma))$. So $\forall x(Rx \supset Px)$ is finally $\mathbf{CIL}^r$-derivable from Γ_9.

Incidentally, we are able to see, in the simple example proof, that $!(\neg Px \vee \neg Qx)$ is $\mathbf{IL}^r_{(0)}$-finally derived at line 14. This, however, is immaterial. Marking proceeds in terms of the insights provided by the proof. So $!(\neg Px \vee \neg Qx)$ counts as a minimal $Dab^{(1)}$-formula in the proof because it was derived at a line that has a subset of $\Omega^{(0)}$ as its condition and that is unmarked in view of $\mathbf{IL}^r_{(0)}$.

Some study of related cases readily leads to the following generalization, which holds for all logics of the **C**-group. The formula $Dab(\Theta)$, derived at line l on the condition Δ in a proof at stage s (i) is a $Dab^{(0)}$-formula at stage s iff $\Theta \subset \Omega^{(0)}$ and $\Delta = \emptyset$, and (ii) is a $Dab^{(i+1)}$-formula at stage s iff $\Theta \subset \Omega^{(i+1)}$, $\Delta \subseteq \Omega^{(i)}$, and line l is unmarked. The general formulation of this follows in Section 6.2.2. From the minimal $Dab^{(i)}$-formulas at stage s of a proof from Γ, 6.2.2 !
one defines $U^{(i)}_s(\Gamma)$ and $\Phi^{(i)}_s(\Gamma)$ in the usual way.

At this point I have illustrated the difference between $\mathbf{HIL}^r$ and $\mathbf{CIL}^r$ and I have clarified the mechanism that is responsible for this difference. The premise set Γ_9 also illustrates the difference between $\mathbf{HIL}^m$ and $\mathbf{CIL}^m$—see Section 6.2.2. 6.2.2 !

The logics of the **C**-group are sequential superpositions of simple adaptive logics. Notwithstanding the complexity of Definition 3.4, the proofs themselves are simple. By not concentrating on the definition, but on the proofs, some insights were found, generalized, and finally proven correct. The insights concern the rules and the marking definition. The rules are simple and identical to those of logics of the **H**-group, except that we do not need the rule RU*. Lines are added by applying all rules of all combining logics—remember that only the conditional rules are different in that they refer to different sets of abnormalities. The sequential character of the combined logic turns out to reduce to sequential marking. The idea is that, at every stage, lines are marked or unmarked in view of the 'innermost' combining adaptive logic (in the sense of Definition 3.4). Relying on this, lines are marked or unmarked in view of the next innermost logic. And so on. Spelled out, the marking definition for the logics of the **C**-group looks as follows.

Definition 3.5.3 *Marking for Reliability: Starting from* $i = 0$*, a line is* i*-marked at stage* s *iff, where* Δ *is its condition,* $\Delta \cap U^{(i)}_s(\Gamma) \neq \emptyset$.

Note that the marking definition interacts with the definition of a $Dab^{(i+1)}$-formula[30] and that $U^{(i+1)}_s(\Gamma)$ is only determined by the set of i-unmarked lines.

Note that this definition does not refer to a specific family of logics, such as the **IL**-family, but holds for all logics of the **C**-group that have Reliability

[30] Although it should be obvious, let me spell it out: a $Dab^{(i+1)}$-formula is a disjunction of members of $\Omega^{(i+1)}$ that is derived on a condition $\Delta \subseteq \Omega^{(i)}$.

6.2.2 ! as their strategy. I again skip the marking definition for minimal Abnormality, postponing it to Section 6.2.2.

The semantics of logics of the **C**-group is rather transparent. Let me illustrate this for Reliability and the **IL**-logics. We start with the **CL**-models of the premise set Γ. These determine the minimal $Dab^{(0)}$-consequences of Γ and so $U^{(0)}(\Gamma)$. The latter determines which of the **CL**-models are $\mathbf{IL}^r_{(0)}$-models of Γ. The so selected models define the $\mathbf{IL}^r_{(0)}$-consequences, which are closed under **CL**, and hence also $U^{(1)}(Cn_{\mathbf{LI}^r_{(0)}}(\Gamma))$. And so on.

By now, I have introduced eighteen different logics of inductive generalization. There is, however, a very good reason to push the matter one step further. The logics of the **C**-group are defined as a superposition of logics combined from simple logics that have increasing sets of abnormalities; $\Omega^{(0)} \subset \Omega^{(1)}$ and so on. That the sets of abnormalities are so connected was required for the logics of the **H**-group. If they were not defined in terms of increasing sets of abnormalities, the union of the consequence sets of the combining logics might be inconsistent. However, the logics of the **C**-group are combined by a superposition. This eliminates the danger for inconsistency because each combining logic turns consistent premise sets into consistent consequence sets. So let us consider superpositions that have disjoint sets of abnormalities. We call their set the **S**-group because they are as it were clean superpositions.

We do not need many new technicalities in order to define these logics. We have already the non-overlapping sets $\Omega^0, \Omega^1, \ldots$ for each of the three families. In terms of these sets of abnormalities, we define, for each $i \in \mathbb{N}$, the logic $\mathbf{LI}^r_i$ as the adaptive logic that is exactly like $\mathbf{LI}^r$ except that it has Ω^i as its set of abnormalities. We define the logics $\mathbf{LI}^m_i$ similarly and do the same for the **IL**-family and for the **G**-family, relating them to their specific Ω^i and to the right strategy. Finally, we define.

$$Cn_{\mathbf{SLI}^r}(\Gamma) = \ldots (Cn_{\mathbf{LI}^r_2}(Cn_{\mathbf{LI}^r_1}(Cn_{\mathbf{LI}^r_0}(\Gamma)))) \ldots \tag{3.5}$$

The logic $\mathbf{SLI}^m$ is defined in a similar way from the $\mathbf{LI}^m_i$ logics. The logics $\mathbf{SIL}^r$, $\mathbf{SIL}^m$, $\mathbf{SG}^r$, and $\mathbf{SG}^m$ are also defined in a similar way, varying the combining logics and their strategy. The marking definitions for the combined logics are as in the case of **C**-group—those logics are also sequential superpositions, just like logics of the **S**-group. The difference between them lies in the fact that $\Omega^{(i)} \subset \Omega^{(i+1)}$ whereas Ω^i and Ω^{i+1} are disjoint.

Of course, I shall now write, for example, Dab^i-formula instead of $Dab^{(i)}$-formula—this change is also required in the generic rule RC. The formula $Dab(\Theta)$, derived at line l on the condition Δ in a proof at stage s (i) is a Dab^0-formula at stage s iff $\Theta \subset \Omega^0$ and $\Delta = \emptyset$, and (ii) is a Dab^{i+1}-formula at stage s iff $\Theta \subset \Omega^{i+1}$, $\Delta \subseteq \Omega^0 \cup \ldots \cup \Omega^i$, and line l is unmarked. From the minimal Dab^i-formulas at stage s of a proof from Γ, one defines $U^i_s(\Gamma)$ and $\Phi^i_s(\Gamma)$ in the usual way.

I now illustrate the difference between the **C**-group and the **S**-group by proofs from $\Gamma_{10} = \{Pa, Qa, \neg Pb \vee \neg Qb\}$. First the $\mathbf{CG}^r$-proof.

1	Pa	premise	$\emptyset$	
2	Qa	premise	$\emptyset$	
3	$\neg Pb \vee \neg Qb$	premise	$\emptyset$	
4	$\forall x Px$	1; RC^0	$\{\pm Px\}$	$\checkmark^6$

5	$\forall x Qx$	2; RC^0	$\{\pm Qx\}$	$\checkmark^6$
6	$\pm Px \vee \pm Qx$	1, 2, 3; RU	$\emptyset$	
7	$\forall x(Px \supset Qx)$	1, 2; RC^1	$\{Px \wedge \pm Qx\}$	$\checkmark^9$
8	$\forall x(Qx \supset Px)$	1, 2; RC^1	$\{Qx \wedge \pm Px\}$	$\checkmark^{11}$
9	$\pm Px \vee (Px \wedge \pm Qx)$	1, 2, 3; RU	$\emptyset$	
10	$Px \wedge \pm Qx$	9; RC^0	$\{\pm Px\}$	$\checkmark^6$
11	$\pm Qx \vee (Qx \wedge \pm Px)$	1, 2, 3; RU	$\emptyset$	

Remember the marking definition. We first mark for degree 0. This means that the marks with superscript 6 are added. So $\pm Px \vee (Px \wedge \pm Qx)$ and $\pm Qx \vee (Qx \wedge \pm Px)$ are minimal $Dab^{(1)}$-formulas, whence lines 7 and 8 are marked.

Line 10 is only inserted to illustrate that $\pm Px \vee (Px \wedge \pm Qx)$ is a minimal Dab^1-consequence of Γ_{10} whereas $Px \wedge \pm Qx$ is not (because line 10 is marked).

Next consider the $\mathbf{SG}^r$-proof, in which I do not repeat the premises.

4	$\forall x Px$	1; RC^0	$\{\pm Px\}$	$\checkmark^6$
5	$\forall x Qx$	2; RC^0	$\{\pm Qx\}$	$\checkmark^6$
6	$\pm Px \vee \pm Qx$	1, 2, 3; RU	$\emptyset$	
7	$\forall x(Px \supset Qx)$	1, 2; RC^1	$\{Px \wedge \pm Qx\}$	
8	$\forall x(Qx \supset Px)$	1, 2; RC^1	$\{Qx \wedge \pm Px\}$	
9	$\pm Px \vee (Px \wedge \pm Qx)$	1, 2, 3; RU	$\emptyset$	
10	$Px \wedge \pm Qx$	1, 5; RU	$\{\pm Px\}$	$\checkmark^6$
11	$\pm Qx \vee (Qx \wedge \pm Px)$	1, 2, 3; RU	$\emptyset$	

The same lines are still marked in view of $\pm Px \vee \pm Qx$, which is a minimal Dab^0-formula. However, $\pm Px \vee (Px \wedge \pm Qx)$ and $\pm Qx \vee (Qx \wedge \pm Px)$ are not Dab^1-formulas because $\pm Px$ and $\pm Qx$ are not members of Ω^1. Moreover, $Px \wedge \pm Qx$ is not a Dab^1-formula either because line 10 is marked. The fact that $\mathbf{G}^r_1$ has Ω^1 as its set of abnormalities, rather than $\Omega^{(1)}$, causes lines 7 and 8 to be unmarked at stage 10 and actually causes $\forall x(Px \supset Qx)$ and $\forall x(Qx \supset Px)$ to be final $\mathbf{SG}^r$-consequences of Γ_{10}.

A warning has to be issued with respect to strategies. Where *simple* adaptive logics were considered, the Minimal Abnormality strategy always offers a richer consequence set than Reliability. This does not hold for combined adaptive logics that are sequential superpositions of simple adaptive logics. If the 'innermost' simple adaptive logic has Minimal Abnormality as its strategy, it may have a larger consequence set than its variant that has Reliability as its strategy. This larger consequence set may contain *Dab*-formulas of the next innermost simple adaptive logic. Some of these may not be derivable by the Reliability strategy.[31] So the strengthening of the innermost logic may cause the consequence set of the next innermost logic to be weaker. A simple example (with thanks to Mathieu Beirlaen) is offered by a premise set we have already seen: $\Gamma_{10} = \{Pa, Qa, \neg Pb \vee \neg Qb\}$. The $\mathbf{SG}^m$-consequence set of Γ_{10} is identical

[31] In Chapter 5 I shall prove Theorem 5.6.2, which states that the derivable *Dab*-formulas are determined by the lower limit logic only. This might confuse the reader who would return to the present chapter. The explanation is simple: every combining logic adds certain generalizations to its consequence set. This consequence set, say S, is the premise set of the next combining adaptive logic, but is also closed under **CL**. Note that S may very well contain *Dab*-formulas (of the *next* combining logic) that are not derivable from the *previous* premise set.

to the **CL**-consequence set of $\Gamma_{10} \cup \{\forall x Px \vee \forall x Qx\}$. This contains, for example, $Pb \vee Qb$, and hence also $(Pb \wedge \neg Qb) \vee (\neg Pb \wedge Qb)$. The $\mathbf{SG}^r$-consequence set of Γ_{10} is identical to the **CL**-consequence set of $\Gamma_{10} \cup \{\forall x(Px \supset Qx), \forall x(Qx \supset Px)\}$. This contains, for example, $\neg Pb \wedge \neg Qb$. I shall consider the proofs in Section 6.2.3.

Section 6.2.3 nazien

This illustrates the difference between $\mathbf{SG}^m$ and $\mathbf{SG}^r$ and shows that their consequence sets are incommensurable for this premise set. $\mathbf{SG}^m$ squeezes out more at the level of the most general hypotheses, but what one obtains is only a disjunction of generalizations as well as disjunctive predictions, even on the object b. $\mathbf{SG}^r$ proceeds more carefully, delivers nothing at all concerning generalizations of degree 0, but delivers more concerning generalizations of degree 1. Even very weak extensions of the premise set cause $\mathbf{SG}^m$ and $\mathbf{SG}^r$ to agree. That is not surprising. Adding a sufficient amount of new data will cause all our logics of inductive generalization to agree, as they should.

It seems wise to insert a table listing the 24 logics of inductive generalization that we met.

	LI-family		**IL**-family		**G**-family	
simple	$\mathbf{LI}^r$	$\mathbf{LI}^m$	$\mathbf{IL}^r$	$\mathbf{IL}^m$	$\mathbf{G}^r$	$\mathbf{G}^m$
H-group	$\mathbf{HLI}^r$	$\mathbf{HLI}^m$	$\mathbf{HIL}^r$	$\mathbf{HIL}^m$	$\mathbf{HG}^r$	$\mathbf{HG}^m$
C-group	$\mathbf{CLI}^r$	$\mathbf{CLI}^m$	$\mathbf{CIL}^r$	$\mathbf{CIL}^m$	$\mathbf{CG}^r$	$\mathbf{CG}^m$
S-group	$\mathbf{SLI}^r$	$\mathbf{SLI}^m$	$\mathbf{SIL}^r$	$\mathbf{SIL}^m$	$\mathbf{SG}^r$	$\mathbf{SG}^m$

The reader became acquainted with ampliative adaptive logics as well as with three kinds of combined adaptive logics. Meanwhile, I have shown that there is a large variety of logics of inductive generalization. Some readers might think that one should simply apply the strongest logic of this set, but this is not always the right choice. It is for example possible that the final consequences assigned to a data set by a combined logic are generalizations of a low degree, or disjunctions of them, whereas the final consequences assigned to the same data set by the corresponding simple logic comprises more interesting generalizations of a higher degree. Much will depend on the knowledge or presumptions a researcher has about a certain domain. Much will also depend on whether the aim of the application is to act on the basis of a prediction, in which case it is wise to be careful, or to set up a set of interesting experiments in view of the present data, in which case very general hypotheses will soon be shown false if false they are.

Before moving on to background knowledge, let me answer a possible complaint. Some people will wonder whether all these logics have any use. They might argue that Reichenbach's Straight Rule from [Rei38] even provides a method for deriving probabilistic hypotheses and so can be downgraded to apply to inductive generalizations as considered here. The idea is that one takes one's past experiential data as representative for all possible data. In other words, one takes the universe to be constituted by facts that are similar to those experienced. One takes them to occur in the universe with the frequencies that are present in our past experience, a matter disregarded in this chapter. As Reichenbach argued, in the end the Straight Rule leads to success if it is possible to obtain success. So are the adaptive logics of inductive generalization not a useless complication?

Obviously, I would not have written this chapter if I did not think to have an answer to that complaint. Go over all the premise sets introduced in this chapter. To almost none of them can the Straight Rule be applied. The Straight Rule can be applied if all properties of all observed objects are known. It can be applied, although a supplementary rule is then involved, if there is a set of objects of which all properties are known and every other observed object has properties that are a subset of the properties in the first set. The involved supplementary rule is that every partially known object agrees with a fully known object. Most premise sets we considered in this chapter require a more complex reasoning.

Where those restrictions are met, all adaptive logics introduced in this chapter lead to exactly the same result as the Straight Rule. To prove this is a bit tedious, but the following insights help. In so restricted cases, the derivable *Dab*-formulas have, for all twenty-four logics, one disjunct only. If that is so, Reliability and Minimal Abnormality lead to the same result. Next, it is not difficult to see that, if the known objects are completely described, the simple adaptive logics of the three families classify the same generalizations either as derivable at a stage and unmarked in all extensions or as not derivable at a stage or marked in an extension. Finally, if the known objects are completely described, none of the combined logics can offer anything more than the simple logic on which it is based. So, indeed, the adaptive logics of this chapter differ from each other only for premise sets that do not meet the restrictions.

So the question is whether premise sets that do not meet the restrictions are interesting. Needless to say, premise sets that meet the restrictions are never available as such. They are only obtained if it is determined beforehand which properties are qualified as relevant, if the set of these properties is sufficiently small, and if the properties in this set are easily observable, either as such or by performing certain experiments. Deciding beforehand which properties are qualified as relevant is the tricky matter here. One may be lucky by selecting, on the basis of background knowledge or of a worldview, a set of 'easily' observable properties that display a regularity. Whenever this is not the case, however, one needs the possibility to broaden one's scope. In order to do so, one needs to be able to assess which other properties might play a role. The Straight Rule has nothing to offer in this respect.

3.6 Handling Background Knowledge

The easiest way to handle background knowledge is by adding it to the available data, in other words to the premise set. Two problems lurk around the corner though. The first is that the background knowledge may be inconsistent, the second that it is falsified by the data. As this chapter is written from a classic outlook, I shall largely disregard the first problem. Tackling it is beyond the scope of a chapter that is meant to introduce ampliative as well as combined adaptive logics, but the reader will nose, in view of the previous chapter, in which way the problem should be tackled.

The possibility of falsification obviously cannot be neglected even within a fully classical framework—no classical logician could locate anything deviant in Popper's work. The trouble is that Popper's advice to discard falsified theories is too simple. For more than thirty years—I mean [Lau77]—everyone who can read and has access to the literature should know that scientists often continue to reason from falsified theories. They typically consider the falsifications as problems but keep relying on other consequences of the falsified theory. In other words, discarding falsified theories is not always a good advice because it would leave one without any information about the domain.

It follows that two different cases need to be distinguished: those in which we discard a falsified theory altogether and those in which we only discard the falsified consequences of the theory. In the first case, I shall say that one handles the theories strictly, in the second case that one handles them pragmatically. Note that a falsification may only be discovered after some reasoning. So we are in defeasible waters here.

A further distinction has to be made. Part of our background knowledge consists of separate statements whereas other parts are theories—for present purposes, any non-singleton set of statements will be considered to be a theory. Handling theories, whether strictly or pragmatically, will turn out to be a trifle more difficult than handling separate statements.

The picture is not complete yet. That the data falsify a piece of background knowledge is the simple case. More often, however, the data falsify a set of background knowledge without falsifying any single theory or any single separate statement in the set. In such cases, it is important to take into account that elements of our background knowledge may differ in plausibility. One will reject the piece of background knowledge that is least plausible anyway. So we need plausibilities or priorities or preferences—I shall take these words to have the same meaning in the present context. Incidentally, this makes the matter even more defeasible. With all this in mind, let us go for it.

Priorities will be expressed within the object language. This has two advantages over the layered tuples of premises that are popular with some computer scientists. The tuples have the form $\langle \Gamma_0, \Gamma_1, \ldots, \Gamma_n \rangle$ in which[32] Γ_0 comprises the certainties (the so-called real premises), Γ_1 comprises the very plausible statements, Γ_2 the less plausible statements, and so on up to Γ_n. The first advantage of expressing the priorities in the object language is that we have to handle a single premise set instead of a tuple. The second advantage is that it results in a richer framework. It enables one, for example, to express that either A has plausibility 1 or B has plausibility 2.

Priorities will be expressed by one or more symbols $\diamond$ if they pertain to statements and theories that are handled strictly and by one or more symbols $\Diamond$ if they pertain to entities that are handled pragmatically. Let A be a statement of the predicative language. That A, in which no diamond occurs, is a premise means that A belongs to the data, to which certainty is attached. That $\Diamond A$, respectively $\diamond A$, is a premise will express that A has the highest priority lower than certainty. That $\Diamond\Diamond A$, respectively $\diamond\diamond A$, is a premise expresses that A has the next highest priority, and so on. To simplify the notation, I shall write $\Diamond^n A$, respectively $\diamond^n A$, to abbreviate A preceded by n diamonds (of the same sort).

Let me at once mention the way to handle a theory $T = \langle \Gamma, \mathbf{CL} \rangle$ that has priority n—given the context, I suppose that all theories are closed under **CL**. To handle T, we let $\{\Diamond^n(A_1 \wedge A_2 \wedge \ldots \wedge A_k) \mid k \geq 1; A_1, A_2, \ldots, A_k \in \Gamma_n\}$ be a subset of the premise set; similarly for $\diamond$ if the theory is handled strictly. If Γ is a finite set of statements, we may just as well put all of these in conjunction

[32]The Γ_i in this expression have obviously nothing to do with the Γ_i I use elsewhere in this chapter to denote specific premise sets.

and prefix $\Diamond^n$ or $\diamond^n$ to this conjunction.

In order to avoid very complex systems in this introductory chapter, I shall presuppose that all pieces of background knowledge, whether separate statements or theories, have a distinct priority. This unrealistic restriction will be removed in Section 6.2.2. At any point in time, there will be at most finitely many pieces of background knowledge. By the simplification just introduced each of them has a specific priority, which is different from the priority of all other pieces of background knowledge. It follows that only finitely many priorities are involved. This completes the description of premise sets that contain background knowledge.

Let us move on to the adaptive logics that handle the priorities. The underlying idea of these logics is to add to the certainties first as much as possible of the highest priority level; to the result of this one adds as much as possible of the next priority level; and so on. From a technical point of view we upgrade the standard language to a modal one by adding the symbols $\Diamond$ and $\diamond$. The meaning of the modalities is fixed by a modal logic: a predicative version of the logic **K**. The semantic metalanguage will contain the set of pseudo-constants $\mathcal{O}$—see Chapter 1 for the reasons to do so. Let $\mathcal{W}_{m\mathcal{O}}$ denote the set of wffs of the modal pseudo-language $\mathcal{L}_{m\mathcal{O}}$.

A **K**-model M is a quintuple $\langle W, w_0, \mathcal{R}, D, v\rangle$ in which W is a set of worlds, $w_0 \in W$, R is a binary relation on W, D a non-empty set and v an assignment function. The assignment function v is as follows:[33]

C1.1 $v\colon \mathcal{W}_{\mathcal{O}} \times W \to \{0,1\}$
C1.2 $v\colon \mathcal{C} \cup \mathcal{O} \times W \to D$ (where $D = \{v(\alpha, w) \mid \alpha \in \mathcal{C} \cup \mathcal{O}; w \in W\}$)
C1.3 $v\colon \mathcal{P}^r \times W \to \wp(D^r)$

The valuation function, $v_M\colon \mathcal{W}_{m\mathcal{O}} \times W \to \{0,1\}$, determined by the model M is defined by:

C2.1 where $A \in \mathcal{S}$, $v_M(A, w) = v(A, w)$
C2.2 $v_M(\pi^r\alpha_1 \ldots \alpha_r, w) = 1$ iff $\langle v(\alpha_1, w), \ldots, v(\alpha_r, w)\rangle \in v(\pi^r, w)$
C2.3 $v_M(\alpha = \beta, w) = 1$ iff $v(\alpha, w) = v(\beta, w)$
C2.4 $v_M(\neg A, w) = 1$ iff $v_M(A, w) = 0$
C2.5 $v_M(A \vee B, w) = 1$ iff $v_M(A, w) = 1$ or $v_M(B, w) = 1$
C2.6 $v_M((\exists\alpha)A(\alpha), w) = 1$ iff $v_M(A(\beta), w) = 1$ for at least one $\beta \in \mathcal{C} \cup \mathcal{O}$
C2.7 $v_M(\Diamond A, w) = 1$ iff $v_M(A, w') = 1$ for at least one w' such that Rww'.
C2.8 $v_M(\diamond A, w) = 1$ iff $v_M(A, w') = 1$ for at least one w' such that Rww'.

All other logical symbols are explicitly defined in the usual way. A **K**-model M verifies $A \in \mathcal{W}_m$, abbreviated as $M \Vdash A$, iff $v_M(A, w_0) = 1$. $\Gamma \vDash_{\mathbf{K}} A$ (A is a semantic consequence of Γ) iff all **K**-models of Γ verify A. $\vDash_{\mathbf{K}} A$ (A is **K**-valid) iff A is verified by all **K**-models.[34]

[33] I refer to the **CL**-semantics for the justification of the peculiarities of the semantic format.

[34] For those familiar with modal logics, a bit of explanation seems useful to relate the style of the semantics to different styles. One may define a function d that assigns to each $w \in W$ its domain $d(w) = \{v(\alpha, w) \mid \alpha \in \mathcal{C} \cup \mathcal{O}\}$. If an element of an r-tuple of $v(\pi^r, w)$ does not belong to $d(w)$, then the r-tuple does not have any effect on the valuation. So $v(\pi^r, w) \in \wp(d(w)^r)$". Note that, if Rww', the question whether $v(\alpha, w)$ is or is not a member of $d(w')$ is immaterial for the value of $v_M(A, w)$. For example, the value of $v_M(\Diamond Pa, w)$ is determined by the values of $v(a, w')$ and $v(P, w')$ for those w' for which Rww'. The semantics may be rephrased as a counterpart semantics: $\mathsf{a} \in d(w)$ is a counterpart of $\mathsf{b} \in d(w')$ just in case there is an $\alpha \in \mathcal{C} \cup \mathcal{O}$ such that $v(\alpha, w) = \mathsf{a}$ and $v(\alpha, w') = \mathsf{b}$. An $\alpha \in \mathcal{C} \cup \mathcal{O}$ may be seen as picking a specific counterpart 'path' on W.

This seems the best point to add some comments on reading of the diamonds as expressing plausibility. The symbols $\lozenge$ and $\diamond$ have exactly the same meaning in **K**. We nevertheless need both of them because they will be handled differently by the adaptive logics.

Some people are confused about the meaning of $\neg\lozenge\neg A$—the other may skip the rest of this paragraph. The confusion is caused by the failure to distinguish between (i) "A is plausible" is not derivable from a given premise set and (ii) "A is not plausible" is derivable from a given premise set. Put differently and in terms of an example, $\lozenge p$ is false in some models of $\{q, \lozenge r\}$, and so is not a semantic consequence of $\{q, \lozenge r\}$, but $\lozenge p$ is true in other models of the premise set, whence $\neg\lozenge p$ is not a semantic consequence of the premise set either. Note also that $\neg\lozenge^2 A$ excludes neither A nor $\lozenge^i A$ for $i \neq 2$. In other words, that A is not plausible to degree 2 does not exclude that it is true, or plausible to any other degree.

To avoid unnecessary complications, I shall not spell out the proof theory of the modal logic **K** but suppose that it is available. So let us turn to the adaptive logic that handles background knowledge in the desired way. It will be a combined adaptive logic, just like the logics from the **C**-group and the **S**-group in Section 3.5. As the reader is already familiar with combined adaptive logics, I shall first present the logic and only thereafter discuss an example of a proof.

What counts as an abnormality is that a formula is plausible (to some degree) and nevertheless false. We have to take two kinds of priorities into account. I recall that $\diamond$ refers to the strict handling of background knowledge and $\lozenge$ to the pragmatic handling. Moreover, that something plausible is false is worse as it is more plausible. So we define a sequence of sets of abnormalities, for every $i \geq 1$,

$$\Omega^i = \{\exists(\lozenge^i A \wedge \neg A) \mid A \in \mathcal{A}\} \cup \{\exists(\diamond^i B \wedge \neg B) \mid B \in \mathcal{W}_s\}\,,$$

plaats nazien

in which $\mathcal{A}$ is the set of (non-modal) atoms from the second paragraph of Section 3.4. There is a very good reason for there to be two kinds of abnormalities. This will become clear in a few paragraphs.

The logic **K** is combined with the sets of abnormalities and with the Reliability strategy, into an infinite sequence of simple adaptive logics $\mathbf{K}^r_i$ $(i \geq 1)$. Finally, we combine all simple $\mathbf{K}^r_i$ into $\mathbf{K}^r$ by a sequential superposition—note the analogy with the definition of $\mathbf{SG}^r$.

$$Cn_{\mathbf{K}^r}(\Gamma) = \ldots(Cn_{\mathbf{K}^r_3}(Cn_{\mathbf{K}^r_2}(Cn_{\mathbf{K}^r_1}(\Gamma))))\ldots \tag{3.6}$$

As the weakest priority will be represented by a finite number of diamonds in any premise set, $\mathbf{K}^r$ reduces to the combination of finitely many combining logics in every application. I leave it to the reader to define the combined logic $\mathbf{K}^m$ similarly.

The dynamic $\mathbf{K}^r$-proofs (and $\mathbf{K}^m$-proofs) are defined by rules and a marking definition. The rules are exactly like those of the logics of the **S**-group in the previous section. As was the case there, the rules of all combining logics may be applied in any order in $\mathbf{K}^r$-proofs.

Other things are also similar to Section 3.5. I shall write RC^i in proofs, thus referring to the specific logic $\mathbf{K}^r_i$ to which the rule belongs. The conditions of the formulas $A_1, \ldots, A_n$ in RU as well as in RC may contain abnormalities from any Ω^j and actually from several Ω^j. The full conditions are carried over by those rules to the condition of B, as the rules indicate.

The marking definition is identical to Definition 3.5.3 but some parameters should be adjusted. A disjunction of members of (the present) Ω^i that is the formula of a line of a proof counts as a Dab^i-formula iff it is derived on a condition $\Delta \subseteq \Omega^1 \cup \ldots \cup \Omega^{i-1}$. $U^i_s(\Gamma)$ comprises the disjuncts of the minimal Dab^i-formulas at stage s of a proof. The only difference between Definition 3.5.3 and the subsequent one is that we start from $i = 1$ and that $i \geq 1$ in the superscript of $U^i_s(\Gamma)$.[35]

Definition 3.6.1 *Marking for Reliability: Starting from $i = 1$, a line is i-marked at stage s iff, where Δ is its condition, $\Delta \cap U^i_s(\Gamma) \neq \emptyset$.*

The marking definition for Minimal Abnormality is once more postponed plaats "once more"
to Chapter 6. So we go straight to the combination of the logic that handles background knowledge with the one that inductively derives generalizations. It is easy to see which combination is needed. Where **L** is the logic handling inductive generalization and the set of data and background knowledge is closed under $\mathbf{K}^r$, the desired combination is $Cn_{\mathbf{L}}(Cn_{\mathbf{K}^r}(\Gamma))$.

Let us move to an example. Toy examples are obviously bound to be artificial. Let the data be $\{Pa, \neg Qa, \neg Pb, Qc, Rc, \neg Qd\}$, let the background knowledge comprise (i) a theory $T_1 = \{\forall x(Px \supset Rx), \forall x(Rx \supset \neg Qx)\}$ that is handled strictly and has priority 1 (single diamond), (ii) a singular statement $\forall x(Px \supset Qx)$ that is handled pragmatically and has priority 2, and (iii) a singular statement $\forall x(Rx \supset Px)$ that is handled strictly and has priority 3. So the premise set will comprise the nine formulas displayed in the following proof:

1	Pa	premise	$\emptyset$
2	$\neg Qa$	premise	$\emptyset$
3	$\neg Pb$	premise	$\emptyset$
4	Qc	premise	$\emptyset$
5	Rc	premise	$\emptyset$
6	$\neg Qd$	premise	$\emptyset$
7	$\diamond\forall x(Px \supset Rx)$	premise	$\emptyset$
8	$\diamond\forall x(Rx \supset \neg Qx)$	premise	$\emptyset$
9	$\diamond(\forall x(Px \supset Rx) \wedge \forall x(Rx \supset \neg Qx))$	premise	$\emptyset$
10	$\Diamond^2\forall x(Px \supset Qx)$	premise	$\emptyset$
11	$\diamond^3\forall x(Rx \supset Px)$	premise	$\emptyset$

As the only theory comprises finitely many non-logical axioms, it is sufficient to list the conjunction of these axioms preceded by the right number of diamonds. So lines 7 and 8 are obviously redundant.

Let us first see what becomes of T_1. To keep the proof within the margins, $^\star\Diamond^i A$ will abbreviate $\exists(\Diamond^i A \wedge \neg A)$ and similarly for $\diamond$.

12	$\forall x(Px \supset Rx)$	7; RC^1	$\{^\star\diamond\forall x(Px \supset Rx)\}$
13	$\forall x(Rx \supset \neg Qx)$	8; RC^1	$\{^\star\diamond\forall x(Rx \supset \neg Qx)\}$
14	$\forall x(Px \supset Rx) \wedge \forall x(Rx \supset \neg Qx)$	9; RC^1	$\{^\star\diamond(\forall x(Px \supset Rx) \wedge \forall x(Rx \supset \neg Qx))\}$

Each of the three derived formulas may be derived on different conditions as well. Doing so or deriving other formulas from 7–9 by RC is rather pointless

[35] The reason for the difference is that Ω^0 was not defined in the present context, viz. five paragraphs ago in the text.

because all of them will be marked. Theory T_1 is obviously falsified by Rc and Qc. So one may continue as follows.

15	$\neg\forall x(Rx \supset \neg Qx)$	4, 5; RU	$\emptyset$
16	$^{\star}\diamond\forall x(Rx \supset \neg Qx)$	8, 15; RU	$\emptyset$

At stage 16, line 13 is marked and will remain marked forever. It is easily seen that $^{\star}\diamond^1(\forall x(Px \supset Rx) \wedge \forall x(Rx \supset \neg Qx))$ is derivable from 9 and 15. As soon as it is derived, line 14 is marked and will remain marked forever.

What about line 13 and what about other lines that may be added to the proof and at which formulas are derived from T_1? The answer is very simple: they will all be marked. They all have a condition that comprises a formula of the form $\diamond^1 A \wedge \neg A$ and all such formulas are provably unreliable. This is as desired. We wanted to handle T_1 in a strict way: either all of it is added to the data, in case T_1 is compatible with the data, or all of it is discarded, in case the data contradict T_1. The strict handling is realized by devising $\mathbf{K}^r$ as a flip-flop logic with respect to the modality $\diamond$.

So let me show that $\mathbf{K}^r$ is a flip-flop with respect to the modality $\diamond$ and that this results in the strict handling of formulas and theories qualified with that modality. Note, first of all, that the discussion concerns a specific composing logic of $\mathbf{K}^r$, namely $\mathbf{K}^r_1$. So let us concentrate on this. If no Dab^1-formula is $\mathbf{K}$-derivable from the premises, no line with a condition $\Delta \subset \Omega^1$ will be marked. So the $\mathbf{K}^r_1$-consequences of the premise set coincide with the consequences delivered by the upper limit logic, which is **Triv**. So I have to show that, if a Dab^1-formula is $\mathbf{K}$-derivable from the premises, then the $\mathbf{K}^r_1$-consequences of the premise set coincide with the consequences delivered by the lower limit logic, which is $\mathbf{K}$.

Consider a proof in which a Dab^1-formula has been derived, say

$$(\diamond A_1 \wedge \neg A_1) \vee \ldots \vee (\diamond A_n \wedge \neg A_n)\,, \qquad (3.7)$$

and in which occurs a line at which B is derived by RC^1 on the condition $\Delta \subseteq \Omega^1$ from a formula of a line with the empty condition. As B is derivable on the condition Δ from the premise set,

$$B \vee Dab^1(\Delta)$$

is a $\mathbf{K}$-consequence of the premise set.[36] From this and (3.7) follows, by propositional **CL**:

$$(\diamond A_1 \wedge (\neg A_1 \wedge B)) \vee \ldots \vee (\diamond A_n \wedge (\neg A_n \wedge B)) \vee Dab^1(\Delta)$$

and from this, again by propositional **CL**:

$$(\diamond A_1 \wedge \neg(A_1 \vee \neg B)) \vee \ldots \vee (\diamond A_n \wedge \neg(A_n \vee \neg B)) \vee Dab^1(\Delta)\,.$$

From this, as $\diamond A_i \vdash_{\mathbf{K}} \diamond(A_i \vee \neg B)$, follows:

$$(\diamond(A_1 \vee \neg B) \wedge \neg(A_1 \vee \neg B)) \vee \ldots \vee (\diamond(A_n \vee \neg B) \wedge \neg(A_n \vee \neg B)) \vee Dab^1(\Delta)\,. \qquad (3.8)$$

If (3.8) is a minimal Dab^1-consequence of the premise set Γ, then $\Delta \subseteq U^1(\Gamma)$. Suppose then that (3.8) is not a minimal Dab^1-consequence of Γ. So there is a

[36] This is proved as Lemma 4.4.1.

minimal Dab^1-consequence D of Γ which contains some but not all disjuncts of (3.8). If some members of Δ are a disjunct of D, then $\Delta \cap U^1(\Gamma) \neq \emptyset$. If no member of Δ is a disjunct of D, then B is **K**-derivable from the premise set.[37]

What we learned is that, in the example proof we are constructing, every line that has a condition of the form $\diamond C \wedge \neg C$ will be marked once the right Dab^1-formulas have been derived (and I have shown the way in which such Dab^1-formulas may be constructed). So let us not derive any further formulas from T_1, but move on to $\Diamond^2\forall x(Px \supset Qx)$. Of course, the generalization $\forall x(Px \supset Qx)$ is also falsified by the data, but it is handled pragmatically and this is what I want to illustrate now.

17	$\forall x(Px \supset Qx)$	10; RC^2	$\{^\star\Diamond^2\forall x(Px \supset Qx)\}$	$\checkmark^{18}$
18	$^\star\Diamond^2\forall x(Px \supset Qx)$	1, 2, 10; RU	$\emptyset$	
19	$\Diamond^2(Pd \supset Qd)$	10; RU	$\emptyset$	
20	$Pd \supset Qd$	19; RC^2	$\{^\star\Diamond^2(Pd \supset Qd)\}$	
21	$\neg Pd$	6, 20; RU	$\{^\star\Diamond^2(Pd \supset Qd)\}$	

So notwithstanding the presence of Pa and $\neg Qa$, which jointly falsify $\forall x(Px \supset Qx)$, one may derive a consequence of $\Diamond^2\forall x(Px \supset Qx)$, for example 19, and obtain a diamond-free formula from this on its 'own' condition, in the example $^\star\Diamond^2(Pd \supset Qd)$. This illustrates nicely what pragmatic handling comes to: a plausible background statement (or theory) is falsified, but one nevertheless relies on non-falsified consequences of it.

The reader may wonder whether the derivability of the abnormality 19 does not result in $U^2(\Gamma) = \Omega^2$—compare this to the fact that I showed that $U^1(\Gamma) = \Omega^1$. Well, while $\mathbf{K}^r$ is a flip-flop with respect to the modality $\diamond$, it is not with respect to the modality $\Diamond$. I cannot prove this at this point, but I can indicate where the difference stems from. If the symbol $\diamond$ in (3.8) is replaced by the symbol $\Diamond$, the result is not a Dab-formula any more—please check again the way in which the Ω^i are defined (some three pages before this one). Note also that the difference is related to the distinct way in which the diamonds occur in the Ω^i, not to the difference in number of diamonds. $\mathbf{T}^r$ is a flip-flop with respect to $\diamond^2$ just as much as with respect to $\diamond$, but not with respect to any $\Diamond^i$.

Finally consider $\diamond^3\forall x(Rx \supset Px)$. The involved generalization is not falsified by the data extended with non-falsified consequences of $\forall x(Px \supset Qx)$. So things will proceed smoothly.

22	$\forall x(Rx \supset Px)$	11; RC^3	$\{^\star\diamond^3\forall x(Rx \supset Px)\}$
23	$\neg Rb$	3, 22; RU	$\{^\star\diamond^3\forall x(Rx \supset Px)\}$
24	Pc	5, 22; RU	$\{^\star\diamond^3\forall x(Rx \supset Px)\}$
25	$Pd \vee \neg Rd$	22; RU	$\{^\star\diamond^3\forall x(Rx \supset Px)\}$
26	$\neg Rd$	21, 25; RU	$\{^\star\Diamond^2(Pd \supset Qd), ^\star\diamond^3\forall x(Rx \supset Px)\}$

All formulas derived at the unmarked lines 18–26 are finally derivable from the premise set (1–11). Of course, we are not home yet. We only derived the useful consequences of the premise set (data plus plausible background knowledge). To this result, we now should apply a logic of inductive generalization. If this logic is $\mathbf{G}^r$, we want the following combined consequence set.

$$Cn_{\mathbf{G}^r}(Cn_{\mathbf{K}^r}(\Gamma)) \tag{3.9}$$

[37] It is then derivable by propositional **CL** from every disjunct of D. Indeed, all of these have the form $\diamond(A_i \vee \neg B) \wedge \neg(A_i \vee \neg B)$, which is equivalent to $\diamond(A_i \vee \neg B) \wedge \neg A_i \wedge B$.

In other words, we first upgrade the data with the background knowledge and then apply the logic of inductive generalization to this. In order to do so, we simply continue within the same proof. Moreover, we may apply the rules all involved logics at any point in the proof. The marking definition is obvious. It is identical to Definition 3.6.1, except that, after the lines have been marked for all the logics combined into $\mathbf{T}^r$, they are marked in view of $\mathbf{G}^r$. Please check that, for any premise set, the marking definition leads only to finitely many 'rounds' of marking and that this remains so if the simple adaptive logic $\mathbf{G}^r$ is replaced, for example by the combined $\mathbf{SG}^r$.

Before moving on to the next section, let me make sure that the reader keeps in mind the way in which *Dab*-formulas are defined in the context of the last example proof. Remember that Dab^1-formulas are disjunctions of members of Ω^1, which is the 'innermost' or 'lowest' set of abnormalities in this case. Dab^2-formulas are disjunctions of members of Ω^2 that are derived at an unmarked line on a condition that is a subset of Ω^1. In the example proof, the only interesting Dab^2-formulas are derived on the empty condition because, as we have seen, all lines that have members of Ω^1 in their condition are marked as soon as the right Dab^1-formulas are derived. Dab^3-formulas are disjunctions of members of Ω^3 that are derived at an unmarked line on a condition that is a subset of $\Omega^1 \cup \Omega^2$. In the example proof, no Dab^3-formulas are derivable, which means that $\forall x(Rx \supset Px)$ is not falsified by the data extended with non-falsified consequences of $\forall x(Px \supset Qx)$. As a result of all this, the logic of inductive generalization is not applied to the original set of data, which is $\{Pa, \neg Qa, \neg Pb, Qc, Rc, \neg Qd\}$, but to $\{Pa, \neg Qa, \neg Pb, \neg Rb, Pc, Qc, Rc, \neg Pd, \neg Qd, \neg Rd\}$.[38]

I do not claim that this enriched set will lead to more generalizations than the original data set. One takes background knowledge into account not because one wants to obtain a richer set of generalizations, but because one wants to build upon the theoretical insights (empirical generalizations and theories) gained by previous generations. This obviously cannot be illustrated by toy examples. Nor is the mechanism suitable in the case of toy examples. The mechanism is only suitable because we have a vast set of data, pertaining to a vast set of predicates. Background knowledge gives one a way to systematize these data, even if some of the generalizations and theories require exceptions, viz. pragmatic handling. Not relying on background knowledge leaves one with a mess of unstructured data. This theme is continued in the next section.

Some readers may wonder that 24 logics of inductive generalization were presented (and further ones are suggested), and only one logic to handle prioritized background knowledge, which actually takes care of the strict as well as of the pragmatic handling of background knowledge. As announced I restricted the many ways in which background knowledge may be handled in order to keep this introductory chapter as simple as possible.

3.7 Conjectures

Ampliative consequences derived by a logic of inductive generalization are justified in view of a logical rationale, which is slightly different for every such

[38] Actually, much more effects of the background knowledge is present in the premise set to which the logic of inductive generalization is applied, but these will have no effect on the *new* generalizations that will be derived.

logic. This rationale is fixed by the choice of the abnormalities and possibly by ordering them in several sets and by the role these sets play in the combined logic. The derived generalizations are conjectures in the sense that the logical rationale is not deduction. The logic is defeasible; it is also non-monotonic in that the derived generalizations may be overruled by future data. In this section we shall deal with very different conjectures, for which there is at best a non-logical justification.

Suppose that, at some stage of a proof, $A \vee B \vee C$ is a minimal *Dab*-formula (at that stage), with A, B, and C abnormalities of the right kind. Each of the three abnormalities may prevent certain generalizations to be derivable. As we learn from Wiśniewski's erotetic logic, see for example [Wiś95, Wiś96], the disjunction $A \vee B \vee C$ invokes the question $?\{A, B, C\}$, in words: Which of A, B, and C is the case? A full answer to this question might be A (or B or C), but even $\neg A$ (or $\neg B$ or $\neg C$) would partially answer the question in that it reduces the question to a simpler one.[39]

The effects of introducing $\neg A$ in the proof are obvious: it enables one to derive $B \vee C$. If $A \vee B \vee C$ is the only *Dab*-formula in which A occurs, A will not be a member of $U_s(\Gamma)$ if $B \vee C$ is derived at stage s. So certain lines, which were marked at previous stages in view of the *Dab*-formula $A \vee B \vee C$, will be unmarked at stage s of the proof. Put differently, by stating that certain abnormalities are false—remember that we are in a classical context—certain generalizations will become finally derivable.

If the person constructing the proof is not clumsy, $\neg A$ is not derivable from the data. There may be several reasons for introducing $\neg A$. These reasons may relate to the state of the discipline, it may also relate to a world-view or to a metaphysical theory one subscribes to. But precisely because $\neg A$ is not derived from the data, it should not be put on the same foot as the data. In other words, $\neg A$ should not be introduced as a certainty, but as a statement that is assigned a certain plausibility but may be defeated by (present or) future experience.

The technical handling of this is obvious. One introduces $\neg A$ with the required priority. In other words one introduces the premise $\diamond^i \neg A$ with the suitable i and handles it in the same way as one handles a piece of background knowledge.[40] So we do not need any new logic in this connection.

The reader might still have the feeling that some hocus pocus is going on. So let me explain the latter explicitly. Consider a proof that contains the following lines, in which A, B, and C are abnormalities of the logic $\mathbf{G}^r$ and $A \vee B \vee C$ is a minimal *Dab*-formula.

⋮

11	$A \vee B \vee C$	...	$\emptyset$
12	$\diamond\neg A$	Prem	$\emptyset$
13	$\neg A$	12; RC	$\{\diamond\neg A \wedge A\}$
14	$B \vee C$	11, 13; Ru	$\{\diamond\neg A \wedge A\}$

The fact that $A \vee B \vee C$ is a minimal *Dab*-formula of $\mathbf{G}^r$ cannot cause lines 13 and 14 to be marked. For one thing, A is not a member of the condition

[39] In technical terms, $?\{A, B, C\}$ together with $\neg A$ implies the question $?\{B, C\}$.

[40] The restriction that all pieces of background knowledge (and formulas denying a certain abnormality) have a different priority was introduced in the present chapter to avoid complications. As promised, the general case will be discussed in Chapter 6.

of these lines. It is a conjunct of a member and that cannot result in the lines being marked. Moreover, in view of (3.9), the marking definition first marks with respect to the $\mathbf{K}^r$-abnormalities, and only thereafter with respect to the $\mathbf{G}^r$-abnormalities. So if lines 13 and 14 are unmarked in view of $\mathbf{K}^r$, $B \vee C$ is a *Dab*-formula of $\mathbf{G}^r$. Incidentally, the only case in which these lines may be marked is where not only $\diamond\neg A$ but also $\diamond\neg B$ and $\diamond\neg C$ are premises, or also where A is a disjunct of another *Dab*-formula (of $\mathbf{G}^r$) and, for every disjunct D of this *Dab*-formula, $\diamond\neg D$ is a premise.

In some cases, one will not introduce $\neg A$, but rather one of the abnormalities, in the example A, B, and C. Of course the abnormality will again be introduced with a certain plausibility. The effect is clear. If, for example, B is introduced as plausible and B is derived on the condition $\{\diamond B \wedge \neg B\}$ on a line that is unmarked in view of $\mathbf{K}^r$, then both A and C will not be members of $U_s(\Gamma)$ from that stage s on (unless when they are also disjuncts of different minimal *Dab*-formulas).[41] As a result, certain generalizations will become derivable.[42]

One of the reasons for introducing some of the abnormalities with a certain plausibility may be the researcher's judgement on the relevance of certain predicates. Suppose that someone takes there to be a lawful relation between P, Q, and R and considers other predicates are irrelevant for this relation. If a minimal *Dab*-formula in a $\mathbf{G}^r$-proof contains a disjunct $Px \wedge \pm Sx$, whereas other disjuncts contain only the predicates judged relevant, then considering $Px \wedge \pm Sx$ as plausible is a way to express the irrelevance of S with respect to the relation sought.

That scientists introduce abnormalities or negations of abnormalities with a certain plausibility partly explains that different research traditions will come up with different theories. Such premises obviously do not release one from gathering new information and from testing the obtained generalizations. However, they allow researchers to arrive at conjectures that may later be modified, refined, or rejected. The existence of different research traditions will take care of the required search for potential falsifiers.

3.8 Some Comments in Conclusion

Many new kinds of adaptive logics were introduced in the present chapter. All logics were ampliative and some were combined. In subsequent chapters, we shall see that all these logics may be organized in a systematic way and that their properties may be studied in a unified way. The aim of this and the preceding chapter was to illustrate the logics and their use, not to study them in a decent way.

A different kind of comment concerns the application of the logics. The reader might have obtained the wrong impression that background knowledge, and especially the kind of conjectures introduced in the previous section, enable one to obtain generalizations that do not rely on empirical information. This impression is wrong and two different aspects of the matter have to be kept apart.

---

[41] Obviously the occurrence of $U_s(\Gamma)$ refers to the logic of inductive generalization which is applied after $\mathbf{K}^r$ is applied.

[42] It is instructive to modify the last example proof to the present case, to see that the matter is fully perspicuous.

The first aspect concerns the specific defeasible character of all logics presented in this chapter: any conclusion may be overruled by future empirical evidence. The situation was very different in the previous chapter. There we considered a given theory that turned out inconsistent. The only problem was to interpret the theory as consistently as possible. In the present chapter, new data are bound to come up. So there is the supplementary problem to make sure that our generalizations and theories will survive future experience. The only way to do so is to test them. Popper is right that one should in the first place perform those tests that will most likely lead to falsifications of our hypotheses.

The very different second aspect is that one needs to formulate hypotheses before testing them, and that one should rely on one's past experience in order to formulate them. This is where data and background knowledge comes in, and were world-views may provide one with further means of selection.

In Section 3.3, I offered some comments on the way in which $\mathbf{LI}^r$ may guide research. More may be said about this and it should obviously be extended to the other logics of inductive generalization. Moreover, one should realize that the intended research is in the first place part of the second aspect mentioned in the previous paragraph. Of course, thus suggested observations and experiments will already provide one with data that may falsify the hypotheses arrived at so far. This does not mean, however, that they will always take care of the first aspect discussed two paragraphs ago. More often than not, further observations and experiments will be needed in view of the obtained generalizations.

Some friends, for example Atocha Aliseda, wondered whether the logics presented in the present chapter were able to explain that a single observation is sufficient to derive certain generalizations and to consider them as well-established, whereas a diversity of observations is required before other generalizations will be considered as established. It seems obvious to me that this difference depends on background knowledge. However, it is difficult to phrase the matter in a first-order context and I promised myself to restrict to first-order languages in the present book.