Adaptive Logics for Abduction and the Explication of Explanation-Seeking Processes

Joke Meheus

Centre for Logic and Philosophy of Science
Ghent University
Blandijnberg 2
9000 Ghent, Belgium
tel: ++ 32 9 264 37 86
tel: ++ 32 9 264 41 87
Joke.Meheus@UGent.be

In this paper, I illustrate the main characteristics of abductive reasoning processes by means of an example from the history of the sciences. The example is taken from the history of chemistry and concerns a very small episode from Lavoisier's struggle with the 'air' obtained from mercury oxide. Eventually, this struggle would lead to the discovery of oxygen. I also show that Lavoisier's reasoning process can be explicated by means of a particular formal logic, namely the adaptive logic $\mathbf{LA}^r$. An important property of $\mathbf{LA}^r$ is that it not only nicely integrates deductive and abductive steps, but that it moreover has a decent proof theory. This proof theory is dynamic, but warrants that the conclusions derived at a given stage are justified in view of the insight in the premises at that stage. Another advantage of the presented logic is that, as compared to other existing systems for abductive reasoning, it is very close to natural reasoning.

1. Introduction

The aim of this paper is twofold. First, I want to illustrate the main characteristics of abductive reasoning processes by means of an example from the history of the sciences. Next, I want to show that such reasoning processes can be explicated by means of a formal logic.

The example is taken from the history of chemistry and concerns a very small episode from Lavoisier's struggle with the 'air' obtained from *mercurius calcinatus per se* (mercury oxide, in modern terminology). Eventually, this struggle would lead to the discovery of oxygen.

Readers who have a romantic view on scientific discovery and creativity should immediately be warned: they will probably be disappointed by the example. It is not a story about a series of genius insights that all of a sudden led to one of the most important discoveries in chemistry. To the contrary, the steps that I shall document on were all very small and quite mundane, and some may even question whether they were rational. However, what makes the example so fascinating is that it is one of those extremely rare occasions where one obtains some insight in the microstructure of reasoning—thanks to some notes that were preserved, it is possible to reconstruct, step by step, the individual inferences that Lavoisier made.

As the example will nicely show, abductive reasoning processes have two important characteristics. The first is that abductive steps are combined with deductive steps. The second is that, partly because of this combination, abductive reasoning processes are *dynamic*. For instance, a conclusion reached on the basis of an abductive step may be withdrawn when its negation is derived by deductive means.

These characteristics confront logicians with the problem to design systems that combine (in a sensible way) deductive and abductive inference rules and that can moreover account for the dynamics involved. In Meheus et al. (2006), it is shown that so-called adaptive logics enable one to solve this problem. Adaptive logics are a specific kind of formal systems that are especially suited for the study of reasoning processes that are non-monotonic and/or dynamic.[1]

The logic presented in Meheus et al. (2006) is called $\mathbf{LA}^r$ and is an ampliative extension of first-order Classical Logic (henceforth **CL**). The logic is intended for the abduction of singular

hypotheses and presupposes that, with respect to a specific application, the set of *explananda* and the set of possible *explanantia* are disjoint (but not necessarily exhaustive). Where an *explanandum* can be explained by different *explanantia*, $\mathbf{LA}^r$ allows only for the abduction of their disjunction.

In this paper, I shall show that the logic $\mathbf{LA}^r$ allows for the explication of Lavoisier's reasoning, and thus provides an insight in the logical structure of his reasoning. I shall also argue that this logical explication enables us to counter the claim that Lavoisier's reasoning was fallacious.

A second warning is in order. I am evidently not claiming that Lavoisier was using an adaptive logic. I am even not claiming that he had any explicit ideas about the logic behind his reasoning—he most likely had not. What I do believe, however, is that Lavoisier had sound logical intuitions, and that the logic $\mathbf{LA}^r$ provides us with a justification for this claim.

2. Explaining the Properties of 'Oxygen'

Early March 1775,[2] Lavoisier conducted a series of experiments on the 'air' that is released when one reduces the red calx of mercury (mercury oxide) to mercury, without the addition of charcoal.[3] As Lavoisier believed that the red calx of mercury was mercury combined with *fixed air*,[4] he was convinced before beginning the experiments that the air he obtained from the reduction of the mercury calx would be fixed air—see Holmes (1985, p. 45). However, when he performed the standard limewater test for fixed air, he found out that the air obtained from the mercury calx did not form a precipitate (as it should have, had it been fixed air).

Lavoisier next turned to the other standard qualitative tests for 'airs'. Also these ruled out that the air was fixed air. As he wrote in his notebook,[5]

> One was first of all curious to test the effect of this air on animals. For this purpose one passed it into a jar into which one introduced a bird. One left the bird inside it for a good half minute, without its appearing to suffer there in the least. Removed from the air, it flew away without having suffered in any way. [Holmes (1985, p. 46)]

This test suggested that the air was very respirable. His experiments with burning candles further confirmed that the air involved was 'better' than common air:

> One repeated the experiment of the candle two times, and in large jars. It is charming. The flame is much larger and much clearer and much more beautiful than in common air, but in color no different from an ordinary flame. [Holmes (1985, p. 47)]

Believing by now that the air was better than common air (in the sense that it better supported respiration and combustion than common air does), Lavoisier performed the *nitrous air test*.[6] This quantitative test was designed by Priestley and was considered by Lavoisier as more reliable than the qualitative tests. During the nitrous air test, one portion of nitrous air was mixed with two portions of the air one wanted to test. The diminution in volume provided an indication of the air one was dealing with. For instance, it was known that common air led to a diminution of one-fifth.

During the nitrous air test, there were actually two kinds of results. On the one hand, there were a number of qualitative results, such as the colour of the vapours that were formed and the rapidity of the effect. These results were available almost immediately after the experiment started. On the other hand, there were the quantitative results concerning the diminution of the volume of the mixed airs. These were available only at the end of the test.

The day that Lavoisier performed the nitrous air test (March 31, 1775), he entered the following note in his notebook:[7]

> One introduced [one part of] nitrous air into two parts of this air. It appeared that the red color of the vapors was more marked, and the effect more rapid than with common air.

There was a diminution of

Lavoisier did not finish this sentence. This indicates that he wrote this note at the beginning of the experiment, when he did not yet know the amount of the diminution. He left a space open, where he could later fill in the amount of the diminution, and wrote below it

so that, according to this operation, one could judge that this air is more perfect to common air.

What happened here? The first qualitative results strengthened Lavoisier's belief that he was dealing with an air that is better than common air (the colour of the vapours was more intense and the effect more rapid that in the case of common air). This belief must have been so strong that he entered it as the conclusion of the entire experiment, even before he knew the quantitative results. From the way in which he wrote the note (leaving open a space to fill in later the outcome of the experiment), one may infer that Lavoisier was fully confident that the quantitative results would confirm the belief with which he started the experiment.

But then something fascinating happened. When he actually obtained the quantitative results, Lavoisier changed his mind about the conclusion. Instead of simply entering the amount of the diminution in the open space, he left the unfinished sentence unfinished, and entered instead all the numerical details (which previously he did not bother to mention):

One employed two measures of this air, each 2.7 cubic inches, making together	5.4
One added nitrous air	2.7
	8.1 cubic inches
The 8.1 cubic inches was reduced almost immediately to 4.42	

Having thus filled the open space, he crossed out his first conclusion (namely that the air was more perfect than common air), and now wrote below it:

That is to say, regarding the portion of nitrous air as probably entirely absorbed, there was one cubic inch, that is to say one-fifth, of the air absorbed.

That is about the proportion of common air.

As is clear from this passage, Lavoisier gave priority to the quantitative results of the nitrous air test, and hence, concluded that the air obtained from mercury calx was simply common air.

Obviously, the story does not end here. The problem remained why the qualitative results indicated that the air was better than common air. Lavoisier never arrived at a satisfactory solution to this problem. It was Priestley who eventually solved the problem by demonstrating, on the basis of a revised version of the nitrous air test, that the air obtained from mercury calx was indeed different from common air.

However, for this paper, the rest of the story is not important. In what follows, I shall concentrate on the elements that led to Lavoisier's note of March 31, 1775.

3. Analysis of Lavoisier's Reasoning

The qualitative tests that Lavoisier performed confronted him with two surprising facts concerning the sample of air obtained from the reduction of mercury calx:

F1 Some bird stayed in a sample of the air under investigation for more than half a minute without suffering.

F2 Some candles burned in a sample of the air under investigation with a larger flame than a normal candle does in a sample of common air.

These facts where surprising in view of Lavoisier's initial expectation that the air under investigation was fixed air (and hence that candles would extinguish in it and birds would die in it). Both facts, however, could easily be explained in view of the following generalizations:

G1 If some air is better than common air, then a bird can stay in it for more than half a minute without suffering.
G2 If some air is better than common air, then a candle burns in it with a larger flame than a normal candle does in common air.

These generalizations were common knowledge at that time and were accepted by Lavoisier. From **F1** and **R1**, one can derive, by abduction:

G2 The air under investigation is better than common air.

The same conclusion also follows abductively from **F2** and **G2**.

This is the conclusion that Lavoisier first entered in his notebook. This conclusion was further confirmed by the qualitative results of the nitrous air test. However, the nitrous air test also led to the following fact:

F3 A sample of the air under investigation led to a reduction of one-fifth in the nitrous air test.

As Lavoisier also accepted the following generalization:

G3 Some air is common air if and only if it leads to a reduction of one-fifth in the nitrous air test.

he was able to derive, *deductively*:

C2 The air under investigation is common air.

This is the conclusion which Lavoisier retained and for which he rejected the earlier conclusion **C1**.

4. Characterisics of Lavoisier's Reasoning Process

The reasoning process analysed in the previous section has several interesting properties. The first is that ampliative steps are combined with deductive steps in one and the same reasoning process. For instance, whereas **C1** is derived by means of abduction, **C2** is derived deductively.

The second property is that the reasoning process is *non-monotonic*: earlier conclusions are rejected in view of new information. As soon as **F3** is added as a new premise, Lavoisier rejects **C1**. This seems reasonable in view of the fact that **C1** was only derived *abductively* and that, from **F3** and **G3**, it follows, *deductively*, that **C1** does not hold true. Thus, ampliative conclusions are reviewed when a 'stronger' derivation to their negation is available.

A final characteristic is that inference rules are validated *contextually*. The abductive inference to **C1** is considered as valid *until* **C2** is derived. At that point, it is no longer considered as a valid inference. Lavoisier makes this clear by striking out its conclusion.

5. Logic-Based Approaches to Abduction

In the next section, I shall briefly present the adaptive logic $\mathbf{LA}^r$ and show that it is suitable for the explication of Lavoisier's reasoning. But first I argue why $\mathbf{LA}^r$ is better suited for this explication than the logic-based approaches to abduction which were developed in Artificial Intelligence.[8]

Within logic-based approaches, abductive inferences are perceived as falling under the following argumentation scheme:

(†) $A \supset B,\ B\ /\ A$

This scheme, which is generally known as *Affirming the Consequent*, is evidently not deductively valid. Hence, as the framework of most logic-based approaches is a deductive one, the above scheme is not implemented directly. Instead, abductive inferences are specified as a kind of 'backward reasoning': given a theory T and an *explanandum* B, find an A such that

(1) $T \cup \{A\} \vdash B$.
(2) $T \nvdash B$
(3) $T \nvdash \neg A$.
(4) $B \nvdash A$.
(5) A is 'minimal'.

The first of these requirements needs no explanation. Also the next two requirements are straightforward: (2) warrants that the *explanandum* B is not explained by the background theory, and (3) that the explanatory hypothesis A is compatible with T.[9] (4) is needed to rule out degenerate cases. For instance, we do not want to abduce B as an explanation for itself. Also, if $T \cup \{A\} \vdash B$, then $T \cup \{A \vee B\} \vdash B$, but we do not want $A \vee B$ as an explanation for B. Cases like this are ruled out by requiring that the truth of the explanatory hypothesis is not warranted by the truth of the *explanandum*—this is what (4) comes to. (5) is related to the fact that, when trying to explain some *explanandum*, one is interested in explanations that are as parsimonious as possible. Hence, in view of $A \supset B \vdash_{\mathrm{CL}} (A \wedge D) \supset B$, one needs to prevent that $A \wedge D$ can be abduced, whenever A can. This can be realized by requiring that the explanatory hypothesis is 'minimal'. This notion of minimality can be defined in different ways—one may, for instance, consider an explanatory hypothesis as minimal if no alternative is available that is logically weaker. However, no matter how it is defined, minimality is a *comparative* notion: whether some explanatory hypothesis A is minimal with respect to some *explanandum* B depends on the available alternatives.

It is important to note that several of the above requirements are *negative*. This does not only hold true for (2)–(4), but also for (5). Indeed, as minimality is a comparative notion, (5) entails:

(5') $T \cup \{C\} \nvdash_{\mathbf{CL}} B$, for every C that satisfies (2)–(4) and in view of which B is not minimal.

One consequence of these negative clauses was already mentioned in the previous section: the consequence relation defined by (1)–(5) is *non-monotonic*. Conclusions that follow abductively from some theory T may be withdrawn when T is extended to $T \cup T'$.

Another consequence is that, at the predicative level, the consequence relation defined by (1)-(5) is not only undecidable, there even is no positive test for it[10]. This is related to the fact that first-order predicate logic is undecidable—if some conclusion A does *not* follow from a set of premises Γ, we may not be able to establish this. Hence, as the consequence relation is partly defined in terms of negative requirements, it immediately follows that, for undecidable fragments, it lacks a positive test. Suppose, for instance, that for some theory T, some *explanandum* B and some sentence A, (1) is satisfied. In that case, it seems reasonable to conclude that A follows

abductively from T. However, if one is unable to establish that also (2)–(5) are satisfied, no reasoning can warrant that this conclusion is not erroneous.

There are different ways to deal with the lack of a positive test. The one usually followed within Artificial Intelligence is to consider only decidable fragments of first-order logic. The rationale behind this is clear: when dealing with decidable fragments, one may be sure that, for arbitrary theories T and *explananda* B, there is an algorithm for (2)–(5), and hence, that a decision method can be designed for "follows abductively from". From the point of view of applications, however, this is an enormous restriction: many interesting theories are undecidable.

An alternative way is to allow that inferences are made, not on the basis of absolute warrants, but on the basis of one's best insights in the premises. When this second option is followed, abductive reasoning processes not only exhibit an *external* form of dynamics (adding new information may lead to the withdrawal of previously derived conclusions), but also an *internal* one (the withdrawal may be caused by merely analysing the premises). Suppose, for instance, that for some theory T, some *explanandum* B, and some sentence A, one established that (1) is satisfied, and one did not establish that one of (2)–(5) is violated. In that case, it seems rational to conclude that A follows abductively from T. This conclusion, however, is provisional. If at a later moment in time, one is able to show that one of the negative requirements is violated (for instance, because one established that $\neg A$ follows from T), A has to be withdrawn as an explanation for B.

There are several arguments in favour of this second option. The first is that unwanted restrictions are avoided: abduction can be defined for *any* first-order theory. A second argument is that the conclusions of abductive reasoning processes are defeasible anyway. Whether the withdrawal of a conclusion is caused by an external factor or an internal one does not seem to be essential. The third, and most important argument is that, even for decidable fragments, it is often unrealistic to require absolute warrants. Even if a decision method is available, reasoners may lack the resources to perform an exhaustive search, and hence, may be forced to act on their present best insights.

The logic $\mathbf{LA}^r$ follows the second option. This has the advantage that, even for undecidable fragments, it enables one to come to justified conclusions. These conclusions are tentative and may later be rejected, but they constitute, given one's insight in the premises at that moment, the best possible estimate of the conclusions that are 'finally derivable' from the premises.[11]

The logic $\mathbf{LA}^r$ has several other advantages. A first one is that (unlike the systems developed within Artificial Intelligence) it has a *proof theory*. As we shall see below, this proof theory is dynamic (conclusions derived at some stage may be rejected at a later stage), but it warrants that the conclusions derived at a given stage are justified in view of the insight in the premises at that stage. This is especially important as, at the predicative level, there is no positive test for abductive reasoning.

Another advantage of the proposed logic is that it is much closer to natural reasoning than the existing systems. As was mentioned in the beginning of this section, abduction is usually viewed as a form of backward reasoning—"find an A that satisfies the requirements (1)–(5)". The search procedure by which this is realized in the existing systems (for instance, some form of linear resolution) is very different from the search procedures of human reasoners. The logic $\mathbf{LA}^r$ treats abduction as a form of 'forward reasoning': it is an ampliative system that directly validates inferences of the form (†).

The third advantage is related to this: unlike what is the case for the AI approaches to abduction, deductive and abductive steps are nicely integrated into a single system. As a consequence, the logic not only enables one to generate explanatory hypotheses, but also to infer predictions on the basis of explanatory hypotheses and the background theory. This is highly important from the point of view of applications. In scientific contexts, for instance, explanatory hypotheses are typically used to derive predictions which, in turn, may lead to a revision of the original hypotheses.

6. An Adaptive Logic for Abduction

The general idea is extremely simple: it is allowed that the predicative version of (†), namely

(‡) $\quad B(\beta)\,,\ (\forall\alpha)(A(\alpha)\supset B(\alpha))\ /\ A(\beta)$

is applied "as much as possible''. For the moment, this ambiguous phrase may be interpreted as "unless and until $(\forall\alpha)(A(\alpha)\supset B(\alpha))\wedge(B(\beta)\wedge\neg A(\beta))$ turns out to be **CL**-derivable from Γ". So, whenever it is **CL**-derivable from Γ that, for some general rule $(\forall\alpha)(A(\alpha)\supset B(\alpha))$ and some explanandum $B(\beta)$, (‡) cannot be applied consistently (because, $\neg A(\beta)$ is **CL**-derivable from Γ), the application of (‡) is overruled. In view of what we have seen in the previous sections, this is exactly what we want.

There is one general restriction, which is needed to obtain a sensible system. Where $\mathcal{W}$ is the set of closed formulas of the standard predicative language, one needs two sets of truth functions of closed primitive formulas,[12] $\mathcal{W}^e$ and $\mathcal{W}^a$, such that no primitive formula occurs in a member of $\mathcal{W}^e$ as well as in a member of $\mathcal{W}^a$. The sets may but need not be combinatorially closed, in other words, they need not contain all subformulas of their members or all truth-functions of these subformulas.

Intuitively, $\mathcal{W}^e$ is the set of *explananda*, formulas that are considered as requiring an explanation, whereas $\mathcal{W}^a$ is the set of *explanantia*, formulas that, if they can be abduced, form potential explanations for the *explananda*. The requirement that no primitive formula occurs in members of both sets can be easily justified with respect to applications. If one tries to abduce an explanation, one has in mind a phenomenon for which an explanation is sought, and the explanation should be logically independent of the explained phenomenon—everyone rejects (even partial) self-explanations.

To save space, expressions of the form $(\forall\alpha)(A(\alpha)\supset B(\alpha))\wedge(B(\beta)\wedge\neg A(\beta))$ will be abbreviated as $[\![B(\beta),\neg A(\beta)]\!]$ and, in line with what is common for adaptive logics, the formula $[\![B(\beta),\neg A(\beta)]\!]$ will be called an "abnormality".[13] As we will see below, it is possible that a disjunction of abnormalities is **CL**-derivable from a set of premises Γ without any of its disjuncts being derivable from it.

$\mathbf{LA}^r$ can be formulated in the standard format from Batens (2007), which greatly simplifies the technical matters. An adaptive logic **AL** is in standard format if it is characterized as a triple consisting of three elements: (i) **LLL**, a compact and monotonic lower limit logic, (ii) Ω, a set of abnormalities that all have the same logical form, and (iii) an adaptive strategy.

The lower limit logic **LLL** determines the part of the adaptive logic **AL** hat is not subject to adaptation. From a proof theoretic point of view, the lower limit logic delineates the rules of inference that hold unexceptionally. From a semantic point of view, the adaptive models of a premise set Γ are a selection of the **LLL**-models of Γ. The lower limit logic of $\mathbf{LA}^r$ is **CL**, and remember that its premise set is $\langle\Gamma,\mathcal{W}^e,\mathcal{W}^a\rangle$.

Abnormalities are formulas that are presupposed to be false, unless and until proven otherwise. Ω comprises all formulas of a certain (possibly restricted) logical form. In the case of $\mathbf{LA}^r$ the restriction will refer to $\mathcal{W}^e$ and $\mathcal{W}^a$. and the set of abnormalities Ω is defined as $\{(\forall\alpha)(A(\alpha)\supset B(\alpha))\wedge(B(\beta)\wedge(\neg A(\beta))\mid A(\beta)\in\mathcal{W}^a;\ B(\beta)\in\mathcal{W}^e;\ \nvdash_{\mathbf{CL}}(\forall\alpha)(A(\alpha)\supset B(\alpha))\}$. In the present extensional framework, $(\forall\alpha)(A(\alpha)\supset B(\alpha))$ can be taken to express that A contains a (sufficient) cause for B—I write "contains" because A may itself be a conjunction and some of its conjuncts may not be required for warranting B. The second conjunct of an abnormality states that the specific sufficient cause $A(\beta)$ for $B(\beta)$ did not occur. The requirement that the generalization $(\forall\alpha)(A(\alpha)\supset B(\alpha))$ is not a **CL**-theorem has to be added in order to prevent that all models would

display abnormalities. However, as this rules out at once cases in which $A(\beta)$ is a contradiction or $B(\beta)$ is a tautology, this requirement is harmless. (Nobody wants to seek an explanation for a tautology and nobody will accept an explanation by *Ex Falso Quodlibet*.) An adaptive logic presupposes that abnormalities are false unless and until proven otherwise. So, the presupposition of $\mathbf{LA}^r$ is that, if an effect did occur, then all its potential causes (in the weak, extensional, sense) did also occur.

The strategy is Reliability. This strategy warrants that, in cases where more than one explanatory hypothesis can be abduced for the same *explanandum*, only their disjunction is derivable by $\mathbf{LA}^r$. It also warrants that, in cases where there are mutually inconsistent explanatory hypotheses, only those explanations are abduced that are *jointly compatible* with the premises. Both cases will be illustrated in the example at the end of this section.[14]

If one adds to the lower limit logic an axiom schema excluding that abnormalities occur, viz. an axiom schema that reduces abnormal premise sets to triviality, one obtains the so-called upper limit logic. The upper limit logic of $\mathbf{LA}^r$ is somewhat unusual as it refers to the sets $\mathcal{W}^e$ and $\mathcal{W}^a$. It is obtained by extending **CL** with the axiom schema $(\forall\alpha)(A(\alpha)\supset B(\alpha))\supset(B(\beta)\supset A(\beta))$ *provided* $B(\beta)\in\mathcal{W}^e$ and $A(\beta)\in\mathcal{W}^a$. It is easily seen that this comes to the requirement that, if the proviso is met, $(\forall\alpha)(A(\alpha)\supset B(\alpha))$ is logically equivalent to $(\forall\alpha)(A(\alpha)\equiv B(\alpha))$. As the upper limit logic is not interesting in itself, I shall not bother to give it a name.

Let us now turn to the proofs. If the deduction rules are formulated in generic format, they are identical for all adaptive logics in standard format. Let Γ contain the (declarative) premises as before, let the notation

$$A \quad \Delta$$

abbreviate that A occurs in the proof on the condition Δ, and let $Dab(\Delta)$ be the disjunction of the members of a finite $\Delta\subset\Omega$. The rules may be phrased as follows:[15]

PREM If $A\in\Gamma$

$$\frac{\begin{array}{ll}\ldots & \ldots\end{array}}{\begin{array}{ll}A & \varnothing\end{array}}$$

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

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

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

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

In addition to the inference rules, also a *marking definition* is needed. The marking definition determines which lines of a proof have to be marked. Formulas that occur on marked lines are no longer considered to be derived in the proof.

I shall say that $Dab(\Delta)$ is a *minimal Dab-formula* at stage s of a proof if, at that stage, $Dab(\Delta)$ occurs in the proof on the empty condition and, for any $\Delta'\subset\Delta$, $Dab(\Delta')$ does not occur

in the proof on the empty condition. Where $Dab(\Delta_1)$, ..., $Dab(\Delta_n)$ are the minimal Dab-formulas at stage s of the proof, $U_s(\langle \Gamma, \mathcal{W}^e, \mathcal{W}^a \rangle) = \Delta_1 \cup \ldots \cup \Delta_n$ is the set of unreliable formulas at stage s. The marking definition for the Reliability Strategy is as follows:

Definition 1 *Line i is marked at stage s iff, where Δ is its condition, $\Delta \cap U_s(\langle \Gamma, \mathcal{W}^e, \mathcal{W}^a \rangle) \neq \varnothing$.*

If $Dab(\Delta)$ is a minimal Dab-formula at stage s of the proof, then, in as far as one knows in view of the proof at this stage, the premises require one of the abnormalities in Δ to be true but do not specify which one is true. The Reliability Strategy considers all of them as unreliable. So the underlying idea is: if the understanding of the premises provided by the present stage of the proof is correct, the formulas occurring at unmarked lines are derivable from the premises, whereas the formulas occurring at marked lines are not.

Apart from the unstable derivability at a stage, one wants a stable kind of derivability, which is called final derivability.

Definition 2 *A is finally derived from $\langle \Gamma, \mathcal{W}^e, \mathcal{W}^a \rangle$ on line i of a proof at stage s iff (i) A is the second element of line i, (ii) line i is not marked at stage s, and (iii) any extension of the proof in which line i is marked may be further extended in such a way that line i is unmarked.*

Definition 3 $\langle \Gamma, \mathcal{W}^e, \mathcal{W}^a \rangle \vdash_{\mathbf{LA}^r} A$ *(A is finally $\mathbf{LA}^r$-derivable from Γ) iff A is finally derived on a line of a $\mathbf{LA}^r$-proof from $\langle \Gamma, \mathcal{W}^e, \mathcal{W}^a \rangle$.*

Remark that these are definitions, and that they are not intended to have a direct computational use.

For the semantics of $\mathbf{LA}^r$, I refer the reader to Meheus et al. (2006). There, it is also shown that the semantics is sound and complete with respect to the dynamic proof theory.

The rest of this section is devoted to an illustration of the proof theory. I shall present a very simple example and not bother too much about technicalities. I shall concentrate on showing (i) that the logic leads to a nice integration of deductive and abductive steps, (ii) that it can handle the dynamics that is typical of abductive reasoning processes, and (iii) that the inference rule which corresponds to abduction is validated contextually.

Suppose that our set of premises Γ consists of the following generalizations

$$(\forall x)(Px \supset Rx),\ (\forall x)(Px \supset Sx),\ (\forall x)(Qx \supset Sx),\ (\forall x)(Qx \supset Tx),\ (\forall x)(\neg Px \supset Tx)$$

and the following data

$$Ra,\ Rb,\ \neg Sb,\ Sc,\ Sd,\ \neg Td,\ Re,\ Te$$

Let $\mathcal{W}^e$ be the set of all singular formulas that are truth-functions of primitive formulas containing the predicates R, S and T, and $\mathcal{W}^a$ the set of all singular formulas that do not contain these predicates.

One way to start a $\mathbf{LA}^r$-proof from Γ is by entering all the premises:

1	$(\forall x)(Px \supset Rx)$	PREM	$\varnothing$
2	$(\forall x)(Px \supset Sx)$	PREM	$\varnothing$
3	$(\forall x)(Qx \supset Sx)$	PREM	$\varnothing$
4	$(\forall x)(Qx \supset Tx)$	PREM	$\varnothing$
5	$(\forall x)(\neg Px \supset Tx)$	PREM	$\varnothing$
6	Ra	PREM	$\varnothing$

7	Rb	PREM	$\varnothing$
8	$\neg Sb$	PREM	$\varnothing$
9	Sc	PREM	$\varnothing$
10	Sd	PREM	$\varnothing$
11	$\neg Td$	PREM	$\varnothing$
12	Re	PREM	$\varnothing$
13	Te	PREM	$\varnothing$

For each of these lines, the third element forms the "justification" for the formula that constitutes the second element. It contains the line numbers of the formulas from which the formula is derived (obviously empty in the case of premises) as well as the name of the rule by means of which the formula is derived (in the above case the premise rule PREM). The sets at the end of each line are the conditions—also these are obviously empty in the case of premises.

We are now in a position to make inferences from the premises. Let us first concentrate on the explanandum Ra. As is easily observed, the first generalization can be used to 'abduce' an explanatory hypothesis for Ra. In an $\mathbf{LA}^r$-proof from Γ, this is done by applying the rule RC:

14	Pa	1,6; RC	$\{[\![Ra, \neg Pa]\!]\}$

RC allows one to add abductive hypotheses to the proof, but only on a certain condition. This condition is represented by the fifth element of the line. Intuitively, line 14 can be read as: Pa is derivable from the formulas on lines 1 and 6, *unless and until* it can no longer be assumed (consistently) that $[\![Ra, \neg Pa]\!]$ is *false*.

Given our present insights in the premises (represented by the formulas that are explicitly written down in the proof), there is no reason to believe that $\neg Pa$ is true, and hence, it is consistent to assume that $[\![Ra, \neg Pa]\!]$ is false. This is why, at this stage of the proof, Pa is considered to be derivable from the premises (in view of line 14). If, at a later stage of the proof, it would turn out that the condition of line 14 is no longer satisfied, then this line will be 'marked' and the formula that occurs on it will no longer be considered to be derived. (The marking of lines will be illustrated below.)

In view of the formula on line 14, the second generalization allows one to infer the prediction Sa; this is done by means of the rule RU:

15	Sa	2,14; RU	$\{[\![Ra, \neg Pa]\!]\}$

RU is a generic rule that allows one to infer all **CL**-consequences: whenever some formula A is **CL**-derivable from a number of formulas $B_1, \ldots, B_n$ that are considered to be derived in the proof at some stage, then, at that stage, A can be added to the proof by means of RU. Note that RU is an *unconditional* rule: unlike RC, it does not lead to the introduction of new conditions. If, however, some of the B_i to which RU is applied are themselves derived on a non-empty condition, then these conditions are conjoined for the conclusion. Thus, as the formula of line 14 is used to derive the formula on line 15, the condition of the former is 'carried over' to the latter. This is obviously as it should be: if, at a later stage in the proof, the conclusion of line 14 is withdrawn because its condition is no longer satisfied, then all formulas that rely on it should also be withdrawn.

This is a first illustration of the way in which abductive steps and deductive steps are integrated. The rule RC allows one to generate new explanatory hypotheses (for instance, the one on line 14), and RU allows one to derive predictions from these.

Let us now turn to the explanandum Rb. As in the previous case, the rule RC enables us to abduce an explanatory hypothesis for Rb (see line 16 below). However, this time, we are also able to infer, by means of RU, the *negation* of our explanatory hypothesis:

16	Pb	1,7; RC	$\{[\![Rb, \neg Pb]\!]\}$
17	$\neg Pb$	2,8; RU	$\varnothing$

Hence, we are able to infer the following abnormality:

18	$[\![Rb, \neg Pb]\!]$	1,7,17; RU	$\varnothing$

At this stage in the proof, the condition of line 16 is no longer satisfied. As a consequence, the conclusion of line 16 is *withdrawn* from the proof. The withdrawal of a conclusion from the proof is recorded by *marking* the line on which the formula occurs. This is how the proof looks like at stage 18 (lines 1 to 15 are as before):

…

16	Pb	1,7; RC	$\{[\![Rb, \neg Pb]\!]\}$ $\surd^{18}$
17	$\neg Pb$	2,8; RU	$\varnothing$
18	$[\![Rb, \neg Pb]\!]$	1,7,17; RU	$\varnothing$

I shall now show what happens when more than one explanatory hypothesis can be abduced for the same explanandum. Have a look at Sc. In view of the relevant generalizations, the proof can be extended as follows:

19	Pc	2,9; RC	$\{[\![Sc, \neg Pc]\!]\}$
20	Qc	3,9; RC	$\{[\![Sc, \neg Qc]\!]\}$

However, as the reader can verify, the following disjunctions of abnormalities are **CL**-derivable from the premises:

21	$[\![Sc, \neg Pc \vee Sc, \neg(Qc \wedge \neg Pc)]\!]$	2,3,9; RU	$\varnothing$
22	$[\![Sc, \neg Qc \vee Sc, \neg(Pc \wedge \neg Qc)]\!]$	2,3,9; RU	$\varnothing$

The formula on line 21 expresses that $[\![Sc, \neg Pc]\!]$ or $[\![Sc, \neg(Qc \wedge \neg Pc)]\!]$ is true. Hence, it cannot be assumed that both disjuncts are false.

In view of such a disjunction of abnormalities, different strategies are possible. The one followed by $\mathbf{LA}^r$ is very cautious. As (at this stage of the proof) it is unclear which one of the two disjuncts is true, both disjuncts are (at this stage of the proof) considered as 'unreliable'. As a result, all formulas that are derived on the assumption that one of these disjuncts is false, are withdrawn. Thus, in our case, the formula on line 19 is withdrawn in view of the formula on line 21. By an analogous reasoning, the formula on line 20 is withdrawn in view of the formula on line 22:

…

19	Pc	2,9; RC	$\{[\![Sc, \neg Pc]\!]\}$ $\surd^{21}$
20	Qc	3,9; RC	$\{[\![Sc, \neg Qc]\!]\}$ $\surd^{22}$
21	$[\![Sc, \neg Pc \vee Sc, \neg(Qc \wedge \neg Pc)]\!]$	2,3,9; RU	$\varnothing$
22	$[\![Sc, \neg Qc \vee Sc, \neg(Pc \wedge \neg Qc)]\!]$	2,3,9; RU	$\varnothing$

A mark may be removed at a later stage. Suppose, for example, that $[\![Sc, \neg(Qc \wedge \neg Pc)]\!]$ is **CL**-derivable from the premises, and is actually derived in the proof. So it would be clear which of

the two disjuncts of the formula of line 21 is true, viz. the second one. As a result, line 19 would not be marked any more (unless $[\![Sc, \neg Pc]\!]$ is a disjunct of another disjunction of abnormalities).

As we have seen, apart from derivability at a stage, a stable notion of derivability is defined, viz. final derivability. Intuitively, a formula is finally derived on line i of a proof iff it is possible to extend the proof in such a way that line i is unmarked and remains unmarked in every further extension of the proof.

In view of the present premises, lines 19 and 20 will remain marked in any extension of the proof. So neither Pc nor Qc is finally derivable from the premises. However, their disjunction $Pc \vee Qc$ is. This can be seen from the following extension of the proof:

23	$(\forall x)((Px \vee Qx) \supset Sx)$	2,3; RU	$\varnothing$
24	$Pc \vee Qc$	9,23; RC	$\{[\![Sc, \neg(Pc \vee Qc)]\!]\}$

As no minimal disjunction of abnormalities is derivable that has $[\![Sc, \neg(Pc \vee Qc)]\!]$ as one of its disjuncts, the formula on line 24 is finally derivable from the premises.

Also for the explanandum Sd the rule RC enables one to derive a disjunction of explanatory hypotheses:

25	$Pd \vee Qd$	2,3,10; RC	$\{[\![Sd, \neg(Pd \vee Qd)]\!]\}$

This time, however, one of the disjuncts can be eliminated by pure deductive means:

26	$\neg Qd$	4,11; RU	$\varnothing$
27	Pd	25,26; RU	$\{[\![Sd, \neg(Pd \vee Qd)]\!]\}$

This again nicely illustrates how $\mathbf{LA}^r$ allows for the integration of deductive and abductive steps.

Let us finally turn to the situation where different explanatory hypotheses are mutually incompatible with the premises. As may be seen from the following extension of the proof, this is the case for the explanatory hypotheses that are abducible for Re and Te:

28	Pe	1,12; RC	$\{[\![Re, \neg Pe]\!]\}$
29	$\neg Pe$	5,13; RC	$\{[\![Te, Pe]\!]\}$

Although both these hypotheses may be entered at some stage in the proof, neither of them is finally derivable from the premises. This is warranted by the following **CL**-derivable disjunction of abnormalities:

30	$[\![Re, \neg Pe]\!] \vee [\![Te, Pe]\!]$	1,5,12,13; RU	$\varnothing$

As soon as the formula on line 30 is added to the proof, lines 28 and 29 are marked—they remain marked in any extension of the proof.

7. Was Lavoisier's Reasoning Rational?

To some it may seem that Lavoisier, in the course of that particular experiment in March 1775, was not reasoning in a rational way. As some may argue, Lavoisier not only committed the fallacy of "affirming the consequent", he moreover jumped to the conclusion on the basis of partial information.

In the absence of formal logics for abduction, this argument carries some weight. It seems good practice to link rationality to "reasoning according to the standards of an appropriate logic''. Hence, if there are doubts about the logicality of a particular inference form, this immediately casts doubts on its rationality.

In the case of abduction, the suspicion seemed justified. Not only is the inference not deductively valid, many examples of purportedly sound abductions seem to rely on a hidden non-formal reasoning. Indeed, the only sensible formal rule behind them seems to lead inevitably to a set of unsound and even inconsistent conclusions. For instance, from the explananda Qa and Ra and the generalizations $(\forall x)(Px \supset Qx)$ and $(\forall x)(\neg Px \supset Rx)$, (‡) enables one to generate both Pa and $\neg Pa$.

In the previous section, we have seen, however, that it is possible to design a formal logic for abduction that is just as rigourous as, say, Classical Logic. What is important about this logic is that it enables one to distinguish between sound and unsound applications of (‡). We have also seen that this distinction is *contextual*: even if abduction should be invalidated with respect to some of the premises, it may be validated with respect to others. Thus, where the generalizations are $(\forall x)(Px \supset Qx)$, $(\forall x)(\neg Px \supset Rx)$ and $(\forall x)(Sx \supset Rx)$, and the *explananda* are Qa and Ra, abduction should be invalidated with respect to the first two premises (neither Pa nor $\neg Pa$ should be abducible), but should be validated with respect to the third premise (Sa should be abducible). This is precisely what the logic $\mathbf{LA}^r$ allows for.

Applied to Lavoisier's premises, the logic $\mathbf{LA}^r$ allows one to derive **C1** from **F1** and **G1** as well as from **F2** and **G2**. Moreover, as long as **F3** is not added, the conclusion **C1** is *finally derivable* from the premise set. It is only when the fact **F3** is added (together with the generalization **G3**) that the inference to **C1** is invalidated. (In an $\mathbf{LA}^r$-proof, this would be expressed by marking the line on which **C1** is derived.)

Some may still have problems with the fact that Lavoisier derived the conclusion *before* he knew all the results to which the experiment would lead. We have seen, however, that the inference to **C1** was not a deductive one, and that it was not treated by Lavoisier as such. He remembered very well that the inference was provisional and he (literally) deleted it as soon as its negation was derived deductively. This is again as one would expect on the basis of $\mathbf{LA}^r$. Moreover, as I argued in Section 4, making provisional inferences is the only sensible way to proceed in cases where there is no positive test.

What is important is that one remembers the conditions under which a provisional judgment should be revised. Abductive inferences should be considered as fallacious if, and only if, these conditions are not remembered.

---

[1] The first logic in this family was designed by Diderik Batens around 1980 and was meant to interpret (possibly) inconsistent theories *as consistently as possible*. Later the notion of an adaptive logic was generalized in different ways (for instance, to capture ampliative forms of reasoning) and a whole variety of adaptive logics was designed—for an overview, see Batens (2007).

[2] The historical context of the example is discussed at length in Holmes (1985); I refer to this work for more details.

[3] As was known already by the Alchemists, liquid mercury can be converted into a red powder (called *mercurius calcinatus per se*) by heating it; by further heating this powder, the mercury can be recovered from it. The latter reduction can be done without the addition of charcoal which posed a problem for the phlogiston theory—see Holmes (1985). for more details. This was one of the reasons why Lavoisier, among others, was interested in the air that is released during the reduction from mercury oxide to mercury.

[4] The term "fixed air'' was introduced by Stephen Hales to refer to ordinary air in a 'fixed' state. The present-day term is *carbonic acid gas.*

[5] This and the following notes were translated from French to English by Larry Holmes. I follow his translation.

[6] In modern terminology, nitrous air is *nitrogen oxide*.

[7] The note was discovered by Larry Holmes. I follow his analysis of the note as presented on p. 47 of Holmes (1985).
[8] In recent years, abductive reasoning has gained an enormous interest in the domain of Artificial Intelligence. At this moment, a large number of systems is available for a variety of application contexts: diagnostic reasoning, text understanding, case-based reasoning, planning , ... For an interesting overview of AI-approaches to abduction, see Gabriele (2000).
[9] A formula A is said to be compatible with a set of premises Γ iff $\Gamma \nvdash \neg A$
[10] Even if A follows abductively from a theory T and an *explanandum* B, there need not exist any finite construction that establishes this.
[11] Roughly speaking, an 'abductive conclusion' A is finally derivable from a theory T if the requirements (1)-(5) are satisfied—see the next section for a precise definition of this notion.
[12] Primitive formulas are those that contain no logical symbols, except possibly for identity.
[13] The term "abnormality'' refers to formulas that overrule the application of some desired inference rule—in our case the abduction scheme ($\ddagger$).
[14] As is illustrated in Meheus et al. (2006); the Reliability Strategy also guarantees that, if the antecedent of some generalization has been arbitrarily strengthened, only sensible explanations are abduced.
[15] The only rule that introduces non-empty conditions is RC. In other words, before RC is applied in a proof, the condition of every line will be $\emptyset$.

Acknowledgements

Research for this paper was supported by subventions from Ghent University and from the Research Foundation—Flanders (FWO - Vlaanderen).

References

Diderik Batens (2007). A universal logic approach to adaptive logics. *Logica Universalis*, 1:221--242.

Paul Gabriele (2000). AI approaches to abduction. In Dov~M. Gabbay and Philippe Smets, editors, *Handbook of Defeasible Reasoning and Uncertainty Management Systems. Volume 4: Abductive Reasoning and Learning*, pages 35--98. Kluwer Academic Publishers, Dordrecht.

Frederic Lawrence Holmes (1985). *Lavoisier and the Chemistry of Life. An Exploration of Scientific Creativity*. The University of Wisconsin Press.

Joke Meheus and Diderik Batens (2006). A formal logic for abductive reasoning. *Logic Journal of the IGPL*, 14:221-236.