Modelling Abduction in Science by means of a Modal Adaptive Logic

Tjerk Gauderis*
Centre for Logic and Philosophy of Science
Ghent University, Blandijnberg 2, 9000 Gent, Belgium
Tjerk.Gauderis@UGent.be

June 21, 2011

Abstract

Scientists confronted with multiple explanatory hypotheses as a result of their abductive inferences, generally want to reason further on the different hypotheses one by one. This paper presents a modal adaptive logic $\mathbf{MLA^s}$ that enables us to model abduction in such a way that the different explanatory hypotheses can be derived individually. This modelling is illustrated with a case study on the different hypotheses on the origin of the Moon.

keywords: abduction - modelling scientific reasoning processes - multiple explanatory hypotheses - adaptive logics - modal logics

1 Introduction

The aim of this paper is to present a new adaptive logic, called $\mathbf{MLA^s}$, that enables us to model abductive reasoning processes. The goal of these processes is to derive possible explanatory hypotheses (*explanantia*) for puzzling phenomena (*explananda*). For that purpose, this logic contains, in addition to deductive inference steps, defeasible reasoning steps based on an argumentation schema known as *Affirming the Consequent* (combined with Universal Instantiation):

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

It is important to mention that by using this schema we restrict our field of application in two ways. Firstly, we consider abduction only in a *strict* sense,

*Research for this paper was supported by project subventions from the Special Research Fund (BOF) of Ghent University. I am indebted to Joke Meheus, Bert Leuridan, Erik Weber and Dagmar Provijn for their helpful comments on earlier drafts. I further want to thank Mathieu Beirlaen and Hans Lycke for their useful suggestions.

which means that the conditional linking explananda and explanantia must be given. In other words, the modelling of any sort of *creative abduction* – in which the conditionals are created – is not within the scope of this paper.[1] Secondly, we opt for a predicate logic. This is so because we use a material implication to model the relation between *explanans* and *explanandum.* As it is well known that $B \vdash A \supset B$, a propositional logic would allow us to derive anything as a hypothesis. In the predicative case, the use of the universal quantifier can avoid this.[2] Moreover, it raises no major problem for modelling real life situations, as the case study illustrates.

Adaptive Logics This logic is constructed by means of the techniques of the adaptive logics programme.[3] The reasons why an adaptive logic is fit for this job are threefold.

Firstly, it allows for a direct implementation of defeasible reasoning steps (*in casu* applications of *Affirming the Consequent*). This makes it possible to construct logical proofs that nicely integrate defeasible (in this case ampliative) and deductive inferences. This corresponds to natural reasoning processes.

Secondly, the formal apparatus of an adaptive logic instructs exactly which formulas would falsify a (defeasible) reasoning step. As these formulas are assumed to be false (for as long as one cannot derive them), they are called *abnormalities* in the adaptive logic literature. So, if one or a combination of these abnormalities is derived in a proof, it instructs in a formal way which defeasible steps cannot be maintained. This possibility to defeat previous reasoning steps mirrors nicely the dynamics that is found in actual human reasoning.

Thirdly, for all adaptive logics in standard format, as the presented logic $\mathbf{MLA^s}$, there are generic proofs for most of the important metatheoretical properties (including soundness and completeness).[4]

The Problem of Multiple Explanatory Hypotheses This is not the first attempt to explicate abductive reasoning by means of an adaptive logic and this result draws on earlier attempts. However, these earlier attempts have not completely dealt with the problem of multiple explanatory hypotheses.

To explain this problem, consider the following example. Suppose we have to explain the puzzling fact Pa while our background knowledge contains both $(\forall x)(Qx \supset Px)$ and $(\forall x)(Rx \supset Px)$. There are two roads that can be taken. Firstly, we can construct a logic in which we can only derive the disjunction $(Qa \vee Ra)$ and not the individual hypotheses Qa and Ra. This road, called *practical* abduction[5] and adequately modelled by the logics $\mathbf{LA^r}$ and $\mathbf{LA^r_s}$,[6] is

[1]For a more elaborate discussion of creative abduction, see Schurz (2008a, p. 212–231).

[2]For example, compare $\vdash B(\beta) \supset (A(\beta) \supset B(\beta))$ with $\nvdash B(\beta) \supset (\forall\alpha)(A(\alpha) \supset B(\alpha))$.

[3]The general characteristics of adaptive logics will be explained in the next section. For a systematic and thorough overview we refer to Batens (2007) or Batens (2011).

[4]An overview of these can be found in Batens (2007).

[5]According to the definition suggested in Meheus and Batens (2006, pp. 224–225) and used in Lycke (2009).

[6]See Meheus and Batens (2006); Meheus (2007, 2010).

suitable for modelling situations in which one has to *act* on the basis of the conclusions. For instance, in the emergency room, a doctor who finds out that two causes can explain the examined symptoms, needs to take appropriate steps based on the fact that they both can be the case.

Secondly, someone with a theoretical perspective (for instance, a scientist or a detective) is interested in finding out which of the hypotheses is the actual explanation. Therefore it is important that he can *abduce* the individual hypotheses Qa and Ra in order to examine them further one by one. Although there exist adaptive logics that model this *theoretical* kind of abduction[7], these logics have a quite complex proof theory. This is because, on the one hand, one has to be able to derive Qa and Ra separately, but on the other hand, one has to prevent the derivation of their conjunction $(Qa \wedge Ra)$, because it seems counterintuitive to take the conjunction of two possible hypotheses as an explanation. Moreover, if the two hypotheses are actually incompatible, it would lead to explosion in a classical context.

Capturing Hypotheses as Logical Possibilities There is actually a more elegant and natural way out of this problem by adding modalities to our language and deriving the hypotheses $\Diamond Qa$ and $\Diamond Ra$. As $(\Diamond Qa \wedge \Diamond Ra)$ does not imply $\Diamond(Qa \wedge Ra)$ in any standard modal logic, the conjunction problem is automatically solved. This new approach, that will be our route, also nicely coincides with the common idea that hypotheses are possibilities. These features make our logic $\mathbf{MLA^s}$ (which stands for the Modal Logic for Abduction) very suitable for the modelling of actual theoretical abductive reasoning processes as the case study will illustrate.

Structure of the paper In the next section, we will first introduce the main characteristics of an adaptive logic in standard format for readers not familiar with the adaptive logics programme. The approach will be general and not limited to abductive contexts. For a more systematic and detailed overview of adaptive logics, we refer to Batens (2011, 2007, 2004). In the third section, we return to our subject and provide the groundwork for our logic by stipulating the deductive framework, i.e. the language schema and the non-defeasible reasoning steps of our logic. The fourth section will introduce in an informal way the defeasible part of our logic with examples that illustrate how this logic fulfills the different desiderata of modelling abductive reasoning contexts. This informal approach is chosen to give more insight in the functioning of our logic. A formal presentation of our logic is presented in the fifth section, while in the sixth section we will use this logic to model a more elaborate example taken from the recent history of science.

[7]See, for instance, Lycke (2009) and another solution of Lycke (presented at an internal meeting).

2 General Characterization of Adaptive Logics

An adaptive logic in standard format is defined by a triple:

(i) A *lower limit logic* (henceforth **LLL**): a reflexive, transitive, monotonic and compact logic that has a characteristic semantics.[8]
(ii) A *set of abnormalities* Ω: a set of **LLL**-contingent formulas characterized by a logical form, or a union of such sets.[9]
(iii) An adaptive *strategy*.

The lower limit logic **LLL** specifies the stable part of the adaptive logic, anything that follows from the premises by **LLL** will never be revoked. Apart from that, it is also possible in an adaptive logic to derive defeasible consequences. These are obtained by assuming that the elements of the set of abnormalities are 'as much as possible' false. The adaptive strategy is needed to specify 'as much as possible'. This will become clear further on.

As stated before, a key advantage of adaptive logics is their *dynamic proof theory* which mirrors human reasoning. This dynamics is possible because a *line* in an adaptive proof has – next to a line number, a formula and a justification – a fourth element, i.e. the *condition*. A condition is a finite subset of the set of abnormalities and specifies which abnormalities need to be assumed to be false for the formula on that line to be derivable.

The inference rules in an adaptive logic reduce to three generic rules. Where Γ is the set of premises, Θ a finite subset of the set of abnormalities Ω, $Dab(\Theta)$ the (classical) disjunction of the abnormalities in Θ, and where

$$A \quad \Delta$$

abbreviates that A occurs in the proof on the condition Δ, the inference rules are given by the generic rules:

[8]Strictly speaking, the standard format for adaptive logics requests that a lower limit logic contains in addition to the **LLL**-operators also the operators of **CL** (Classical Logic). However, these operators have merely a technical role (in the generic meta-theory for adaptive logics) and are not used in the applications that are presented in this paper. Hence, given the introductory nature of this section, we will not go into further detail. In the logics presented in this paper, this condition is implicitly assumed to be satisfied.

[9]A **LLL**-*contingent* formula is a formula that is not a theorem of **LLL**.

PREM If $A \in \Gamma$:

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

RU If $A_1, ..., A_n \vdash_{\mathbf{LLL}} B$:

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

RC If $A_1, ..., A_n \vdash_{\mathbf{LLL}} B \vee Dab(\Theta)$

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

The premise rule PREM states that a premise may be introduced at any line of a proof on the empty condition. The unconditional inference rule RU states that, if $A_1, \ldots, A_n \vdash_{\mathbf{LLL}} B$ and $A_1, \ldots, A_n$ occur in the proof on the conditions $\Delta_1, \ldots, \Delta_n$, we may add B on the condition $\Delta_1 \cup \ldots \cup \Delta_n$. The strength of an adaptive logic comes with the third rule, the conditional inference rule RC, which works analogously to RU, but introduces new conditions. So, it allows to take defeasible steps based on the assumption that the abnormalities are false. Several examples of how these rules are employed in actual proofs can be found in section 4.

The only thing we still need is a criterion that defines when we consider a line of the proof defeated. At first sight, it seems straightforward to mark[10] lines of which one of the elements of the condition is *unconditionally*[11] derived from the premises. But this strategy, called the *simple strategy*, has usually a serious flaw. If it is possible to derive unconditionally a disjunction of abnormalities $Dab(\Delta)$ that is *minimal*, i.e. there is no $\Delta' \subset \Delta$ such that $Dab(\Delta')$ is unconditionally derived, the simple strategy would ignore this information. This is problematic though because at least one of the disjuncts of the ignored disjunction has to be true. Therefore, more advanced strategies have been developed. The best-known of these are *reliability* and *minimal abnormality*. We can only use the simple strategy in cases that

$$\Gamma \vdash_{\mathbf{LLL}} Dab(\Delta) \text{ only if there is an } A \in \Delta \text{ such that } \Gamma \vdash_{\mathbf{LLL}} A$$

with $Dab(\Delta)$ any disjunction of abnormalities out of Ω. The strategy of an adaptive logic in standard format is indicated by the superscript in the name of the logic.

[10] Defeated lines in a proof are marked instead of deleted, because in general it is possible that they later become unmarked in an extension of the proof.

[11] *Unconditionally* derived is to be understood as derived on the empty condition.

3 The Deductive Frame

Formal Language Schema Let $\mathcal{L}$ be the standard predicative language of **CL** with logical symbols $\neg, \supset, \wedge, \vee, \equiv, \forall$ and $\exists$. We will further use $\mathcal{C}$, $\mathcal{V}$, $\mathcal{F}$ and $\mathcal{W}$ to refer respectively to the sets of individual constants, individual variables, all (well-formed) formulas of $\mathcal{L}$ and the closed (well-formed) formulas of $\mathcal{L}$.

$\mathcal{L}_M$, the language of our logic, is $\mathcal{L}$ extended with the modal operator $\Box$. $\mathcal{W}_M$, the set of closed formulas of $\mathcal{L}_M$ is the smallest set that satisfies the following conditions:

1. if $A \in \mathcal{W}$, then A, $\Box A \in \mathcal{W}_M$
2. if $A \in \mathcal{W}_M$, then $\neg A \in \mathcal{W}_M$
3. if $A, B \in \mathcal{W}_M$, then $A \wedge B, A \vee B, A \supset B, A \equiv B \in \mathcal{W}_M$

It is important to notice that there are no occurrences of modal operators within the scope of another modal operator or a quantifier. We further define the set $\mathcal{W}_\Gamma$, the subset of $\mathcal{W}_M$, the elements of which can act as premises in our logic, as:

$$\mathcal{W}_\Gamma = \{\Box A \mid A \in \mathcal{W}\}$$

It is easily seen that $\mathcal{W}_\Gamma \subset \mathcal{W}_M$.

Lower Limit Logic The **LLL** will be the predicative version of **D**, restricted by the language schema $\mathcal{W}_M$. **D** is characterized by a full axiomatization of predicate **CL** together with two axioms, an inference rule and a definition:

K $\Box(A \supset B) \supset (\Box A \supset \Box B)$
D $\Box A \supset \neg\Box\neg A$
NEC if $\vdash A$, then $\vdash \Box A$
$\Diamond_{\mathbf{df}}$ $\Diamond A =_{df} \neg\Box\neg A$

This logic is one of the weakest normal modal logics that exist and is obtained by adding the **D**-axiom to the axiomatization of the better-known minimal normal modal logic **K**.

The semantics for this logic can be expressed by a standard possible world Kripke semantics where the accessibility relation R between possible worlds is *serial*, i.e. for every world w in our model, there is at least one world w' in our model such that Rww'.

Intended Interpretation As indicated in the introduction, explanatory hypotheses – the results of abductive inferences – will be represented by formulas of the form $\Diamond A$ ($A \in \mathcal{W}$). We will use formulas of the form $\Box B$ to represent explananda, other observational data and relevant background knowledge. Otherwise, this information would not be able to revoke hypotheses as, for instance, A and $\Diamond\neg A$ are not contradictory. The reason why we choose **D** instead of **K** is that we assume that the explananda and background information are consistent. This assumption is modelled by the **D**-axiom.[12]

4 Informal Presentation of $\mathbf{MLA^s}$

Abductive Contexts and the Set of Abnormalities In specifying our set of abnormalities and the strategy, we have to check whether they allow us to model abductive reasoning according to our expectations.

Apart from the fact that by means of this logic we should be able to derive hypotheses according to the schema of *Affirming the Consequent*, we have to make sure that we cannot derive – as a side effect – random hypotheses which are not related to the explanandum. In addition, it is quite straightforward to expect that a logic for hypothesis formation can handle contradictory hypotheses. Finally, as we pointed out in our introduction, it is a nice feature of adaptive logics that they enable us to integrate defeasible and deductive steps. Therefore, we can require that our logic can handle further predictions (based on earlier derived hypotheses) and evidence or counterevidence for them in a natural way.

Since the final form of the abnormalities is quite complex – although the idea behind it is straightforward – we will first consider two more basic proposals that are constitutive for the final form and show why they are unsufficient. Obviously, only closed well-formed formulas can be an element of any set of abnormalities. This will not be explicitly stated each time.

First proposal Ω_1 This first proposal is a modal version of the set of abnormalities of the logic $\mathbf{LA^r_s}$.[13] In this and the further definitions, the metavariables A and B represent (well-formed) formulas, α a variable and β a constant of the language $\mathcal{L}$.

$$\begin{aligned}\Omega_1 \;=\; &\{\Box((\forall\alpha)(A(\alpha) \supset B(\alpha)) \wedge (B(\beta) \wedge \neg A(\beta))) \mid \\ &\text{No predicate that occurs in } B \text{ occurs in } A\}\end{aligned}$$

This means that a derived hypothesis will be defeated if one shows explicitly that the hypothesis cannot be the case. The second line in the definition is to prevent self-explanatory hypotheses.

Simple Strategy For this logic we can use the *simple strategy* which means, as stated before, that we will mark lines for which one of the elements of

[12]For instance, the premiseset $\{\Box\neg Pa, \Box(\forall x)Px\}$ is a set modelling an inconsistent set of background knowledge and observanda. However, in the logic **K**, this set would not be considered inconsistent, because we cannot derive anything from this set by *Ex Falso Quodlibet*. To be able to do this, we need the **D**-axiom.

[13]As proposed in Meheus (2010).

the condition is unconditionally derived. We can easily see that the condition for use of the simple strategy, i.e.

$$\Gamma \vdash_{\mathbf{LLL}} Dab(\Delta) \text{ only if there is an } A \in \Delta \text{ such that } \Gamma \vdash_{\mathbf{LLL}} A,$$

is fulfilled here. Since all premises have the form $\Box A$, the only option to derive a disjunction of abnormalities would be to apply addition, i.e. derive $(\Box A \vee \Box B)$ from $\Box A$ (or $\Box B$), because it is well-known that $\Box(A \vee B) \nvdash \Box A \vee \Box B$ in any standard modal logic.[14]

Contradictory hypotheses The following example shows us that this logic allows us to derive hypotheses according to the schema of *Affirming the Consequent* and is able to handle contradictory hypotheses without causing explosion.

1	$\Box(\forall x)(Px \supset Qx)$	-;PREM	$\emptyset$
2	$\Box(\forall x)(\neg Px \supset Rx)$	-;PREM	$\emptyset$
3	$\Box Qa$	-;PREM	$\emptyset$
4	$\Box Ra$	-;PREM	$\emptyset$
5	$\Diamond Pa$	1,3;RC	$\{\Box((\forall x)(Px \supset Qx) \wedge (Qa \wedge \neg Pa))\}$
6	$\Diamond\neg Pa$	2,4;RC	$\{\Box((\forall x)(\neg Px \supset Rx) \wedge (Ra \wedge \neg\neg Pa))\}$
7	$\Diamond Pa \wedge \Diamond\neg Pa$	5,6;RU	$\{\Box((\forall x)(Px \supset Qx) \wedge (Qa \wedge \neg Pa)),$ $\Box((\forall x)(\neg Px \supset Rx) \wedge (Ra \wedge \neg\neg Pa))\}$

$\Diamond Pa$ and $\Diamond\neg Pa$ are both derivable hypotheses because the conditions on line 5-7 are not unconditionally derivable from the premise set. It is also interesting to note that, because of the properties of the lower limit **D**, it is not possible to derive from these premises that $\Diamond(Pa \wedge \neg Pa)$. The conjunction of two hypotheses is never considered as a hypothesis itself, unless there is further background information that links the two hypotheses in some way.

Predictions and Evidence Suppose that we can extend our premise set with an additional implication.[15] Then, we can continue our example to show that our logic can handle further predictions and (counter)evidence for them in a natural way:

8	$\Box(\forall x)(Px \supset Sx)$	-;PREM	$\emptyset$
9	$\Diamond Sa$	5,8;RU	$\{\Box((\forall x)(Px \supset Qx) \wedge (Qa \wedge \neg Pa))\}$

With the extra implication we can derive the prediction $\Diamond Sa$. As long as we don't have any further information about this prediction (by, for instance,

[14]Strictly speaking, it is also possible to derive a disjunction from the premises by using the **K**-axiom. For instance, $\Box(A \supset B) \vdash \neg\Box A \vee \Box B$, but the first disjunct will always be a possibility and can, hence, not be an abnormality.

[15]Strictly speaking, this is not what we actually do. What we actually do is start a new proof with another premise set (the extended set). But it is easily seen that we can start this new proof with exactly the same lines as the old proof. This way, it looks as if we extended the old proof. This qualification needs to be considered each time we speak about "adding premises" in this paper.

observation), it remains a hypothesis derived on the same condition as $\Diamond Pa$. If we would test this prediction, we would have two possibilities. On the one hand, if the prediction turns out to be false, the premise $\Box\neg Sa$ could be added to our premise set. In this case, we can subsequently derive $\Box\neg Pa$, which would falsify the hypothesis $\Diamond Pa$.

⋮	⋮	⋮	⋮	
5	$\Diamond Pa$	1,3;RC	$\{\Box((\forall x)(Px \supset Qx) \wedge (Qa \wedge \neg Pa))\}$	$\checkmark^{12}$
⋮	⋮	⋮	⋮	
10	$\Box\neg Sa$	PREM	$\emptyset$	
11	$\Box\neg Pa$	8,10;RU	$\emptyset$	
12	$\Box((\forall x)(Px \supset Qx) \wedge (Qa \wedge \neg Pa))$	1,3,11;RU	$\emptyset$	

On the other hand, if the prediction turns out to be true, we could add the premise $\Box Sa$ but this extension of our premise set would not allow us to derive $\Box Pa$. Since true predictions only *corroborate* the hypothesis and do not *prove* it, while false predictions directly *falsify* the hypothesis, one can say that this logic handles predictions in a *Popperian* way.[16]

Contradictions One way a logic of abduction can generate random hypotheses as a side effect, is by allowing to abduce contradictions. How this is possible and how our logic prevents it is illustrated in the following example.

1	$\Box Qa$	-;PREM	$\emptyset$	
2	$\Box(\forall x)((Px \wedge \neg Px) \supset Qx)$	-;RU	$\emptyset$	
3	$\Diamond(Pa \wedge \neg Pa)$	1,2;RC	$\{\Box((\forall x)((Px \wedge \neg Px) \supset Qx) \wedge (Qa \wedge \neg(Pa \wedge \neg Pa)))\}$	$\checkmark^{4}$
4	$\Box((\forall x)((Px \wedge \neg Px) \supset Qx) \wedge (Qa \wedge (\neg Pa \vee Pa)))$	1;RU	$\emptyset$	

Tautologies Still, there are other ways to derive random hypotheses that are not prevented by our first proposal for our set of abnormalities Ω_1. For instance, Ω_1 does not prevent that random hypotheses can be derived from a tautology.

1	$\Box(Qa \vee \neg Qa)$	-;RU	$\emptyset$
2	$\Box(\forall x)(Px \supset (Qx \vee \neg Qx))$	-;RU	$\emptyset$
3	$\Diamond Pa$	1,2;RC	$\{\Box((\forall x)(Px \supset (Qx \vee \neg Qx)) \wedge ((Qa \vee \neg Qa) \wedge \neg Pa))\}$

Therefore, let us adjust the set of abnormalities to obtain the second proposal Ω_2.

[16]We have to remember we devised a logic for modelling abduction and the handling of explanatory hypotheses, not a formal methodology of science. This logic has nothing to say about confirmation of theories.

Second proposal Ω_2 No hypothesis can be abduced from a tautology if the abnormalities have the following form:

$$\begin{aligned}\Omega_2 \quad = \quad & \{\Box((\forall\alpha)(A(\alpha) \supset B(\alpha)) \wedge (B(\beta) \wedge \neg A(\beta))) \\ & \vee\Box(\forall\alpha)B(\alpha) \mid \\ & \text{No predicate that occurs in } B \text{ occurs in } A\}\end{aligned}$$

It is clear that we can keep using the simple strategy with this new set of abnormalities. It is also easily seen that all examples above still hold. Each time we can derive an abnormality of Ω_1, we can derive the corresponding abnormality of Ω_2 by a simple application of addition. Finally, the problem raised by the tautologies, as illustrated in the previous example, is solved in an elegant way, because the form of the abnormalities makes sure that the abnormality will always be a theorem in case the explanandum is a theorem. So, nothing can be abduced from tautologies.

Most parsimonious *explanantia* Still, there is another way to derive random hypotheses that cannot be prevented by Ω_2. Consider, for instance, the following proof.

1	$\Box Ra$	-;PREM	$\emptyset$
2	$\Box(\forall x)(Px \supset Rx)$	-;PREM	$\emptyset$
3	$\Box(\forall x)((Px \wedge Qx) \supset Rx)$	2;RU	$\emptyset$
4	$\Diamond(Pa \wedge Qa)$	1,3;RC	$\{\Box((\forall x)((Px \wedge Qx) \supset Rx) \wedge (Ra \wedge \neg(Pa \wedge Qa))) \vee \Box(\forall x)Rx\}$
5	$\Diamond Qa$	4;RU	$\{\Box((\forall x)((Px \wedge Qx) \supset Rx) \wedge (Ra \wedge \neg(Pa \wedge Qa))) \vee \Box(\forall x)Rx\}$

The reason why we can derive the random hypothesis Qa, is the absence of a mechanism to make sure that the abduced hypothesis is the most parsimonious one and not the result of *strengthening the antecedent* of an implication. Before defining the final and actual set of abnormalities that also prevents this way of generating random hypotheses, we have to introduce a new notation to keep things as perspicuous as possible.

Notation Suppose $A_{PCN}(\alpha)$ is the *prenex conjunctive normal* form of $A(\alpha)$. This is the equivalent form of $A(\alpha)$ where all quantifiers are first moved to the front of the expression and where, consequently, the remaining (quantifier-free) expression is written in conjunctive normal form, i.e. as a conjunction of disjunctions of literals.

$$\begin{aligned}& A_{PCN}(\alpha) = (Q_1\gamma_1)\ldots(Q_m\gamma_m)(A_1(\alpha) \wedge \ldots \wedge A_n(\alpha)) \\ & \text{and} \vdash A_{PCN}(\alpha) \equiv A(\alpha)\end{aligned}$$

with $m \geqslant 0$,$n \geqslant 1$,$Q_i \in \{\forall, \exists\}$ for $i \leqslant m$, $\gamma_i \in \mathcal{V}$ for $i \leqslant m$, $\alpha \in \mathcal{V}$ and $A_i(\alpha)$ disjunctions of literals in $\mathcal{F}$ for $i \leqslant n$.

Then, we can introduce a new notation $A_i^{-1}(\alpha)$ ($1 \leqslant i \leqslant n$) so that we have a way to take out all of the conjuncts of a formula (in PCN form) one by one. In cases where the conjunction consists of only one conjunct (and there is, obviously, no possibility to have a more parsimonious explanation), the substitution with a random tautology will make sure that the condition for parsimony, added in the next set of abnormalities, is satisfied trivially.

$$\begin{array}{lcl} \text{if } n > 1 & : & A_i^{-1}(\alpha) =_{df} (Q_1\gamma_1)\dots(Q_m\gamma_m)(A_1(\alpha) \wedge \dots \wedge A_{i-1}(\alpha) \wedge \\ & & \quad A_{i+1}(\alpha) \wedge \dots \wedge A_n(\alpha)) \\ \text{if } n = 1 & : & A_1^{-1}(\alpha) =_{df} \top \end{array}$$

with $\top$ any tautology of the **LLL**.

Final proposal Ω With this notation we can write the logical form of the set of abnormalities Ω of our logic $\mathbf{MLA^s}$.

$$\begin{array}{rcl} \Omega & = & \{\Box((\forall\alpha)(A(\alpha) \supset B(\alpha)) \wedge (B(\beta) \wedge \neg A(\beta))) \\ & & \vee \Box(\forall\alpha)B(\alpha) \vee \bigvee_{i=1}^{n} \Box(\forall\alpha)(A_i^{-1}(\alpha) \supset B(\alpha)) \mid \\ & & \text{No predicate that occurs in } B \text{ occurs in } A\} \end{array}$$

This form might look complex, but its functioning is quite straightforward. We actually constructed the disjunction of the three reasons why we stop considering $A(\beta)$ as a good explanatory hypothesis for the phenomenon $B(\beta)$, even if we have $(\forall\alpha)(A(\alpha) \supset B(\alpha))$. The disjunction will make sure that the hypothesis is rejected as soon as one of the following three is the case: (i) when $\neg A(\beta)$ is derived, (ii) when $B(\beta)$ is a tautology (and obviously, does not need an explanatory hypothesis) or (iii) when $A(\beta)$ has a redundant part and is therefore not an adequate explanatory hypothesis.

From now on, we will unambiguously shorten this logical form of the abnormalities as

$$!A(\beta) \triangleright B(\beta)$$

which could be read as "$A(\beta)$ is not a valid hypothesis for $B(\beta)$". For the same reasons as stated at the description of Ω_2, we can keep using the simple strategy and all previous examples are still valid.

Example For instance, let's have a look at how our new set of abnormalities solves the previous problem. To make things more clear, the condition is once written out fully.

1	$\Box Ra$	-;PREM	$\emptyset$
2	$\Box(\forall x)(Px \supset Rx)$	-;PREM	$\emptyset$

3	$\Box(\forall x)((Px \wedge Qx) \supset Rx)$	2;RU	$\emptyset$	
4	$\Diamond(Pa \wedge Qa)$	1,3;RC	$\{\Box((\forall x)((Px \wedge Qx) \supset Rx) \wedge$ $(Ra \wedge \neg(Pa \wedge Qa))) \vee \Box(\forall x)Rx$ $\vee\Box(\forall x)(Px \supset Rx)$ $\vee\Box(\forall x)(Qx \supset Rx)\}$	$\checkmark^5$
5	$!(Pa \wedge Qa) \triangleright Ra$	2; RU	$\emptyset$	

5 Formal Presentation of MLA$^{\mathbf{s}}$

We can now present the logic **MLA**$^{\mathbf{s}}$ in a formally precise way.[17] As any adaptive logic in standard format, the logic **MLA**$^{\mathbf{s}}$ is characterized by the triple of a lower limit logic, a set of abnormalities and an adaptive strategy. In this case, the lower limit logic is **D**, the strategy is the simple strategy and the set of abnormalities Ω is, relying on the previously introduced abbreviation, defined by

$$\Omega \quad = \quad \{!A(\beta) \triangleright B(\beta) \mid \text{No predicate that occurs in } B \text{ occurs in } A\}$$

Proof Theory The proof theory is characterized by the three generic inference rules introduced in section 2 and the following definitions.

Definition 1 (**Marking for the simple strategy**). *Line i with condition Δ is* marked for the simple strategy *at stage s of a proof,*[18] *if stage s contains a line of which an $A \in \Delta$ is the formula and $\emptyset$ the condition.*

Definition 2. *A formula A is* derived from Γ at stage *s of a proof iff A is the formula of a line that is unmarked at stage s.*

Definition 3. *A formula A is* finally derived from Γ at stage *s of a proof iff A is derived at line i, line i is not marked at stage s and line i remains unmarked in every extension of the proof.*[19]

Definition 4 (**Final Derivability**). *For $\Gamma \subset \mathcal{W}_\Gamma$: $\Gamma \vdash_{\mathbf{MLA^s}} A$ (A* is finally **MLA**$^{\mathbf{s}}$-derivable from Γ*) iff A is finally derived in a* **MLA**$^{\mathbf{s}}$*-proof from Γ.*

Semantics The semantics of an adaptive logic is obtained by a selection on the models of the lower limit logic. With the simple strategy we consider, for instance, only those models that verify the abnormalities that are derivable (by means of the lower limit logic).

[17] This section is limited to what we need to present this specific logic. For a more general formal presentation of adaptive logics in standard format, we refer to Batens (2011).

[18] A *stage of a proof* is a sequence of lines and a proof is a chain of stages. Every proof starts off with the first stage which is an empty sequence. Each time a line is added to the proof by applying one of the inference rules, the proof comes to its next stage, which is the sequence of lines written so far extended with the new line.

[19] This definition is slightly different from the more general definition mentioned in Batens (2011) because, using the simple strategy, it is not possible that a marked line becomes unmarked at a later stage of a proof.

Definition 5. *A* **D**-*model* M *of the premise set* Γ *is* simply all right *iff* $\{A \in \Omega \mid M \models_{\mathbf{D}} A\} = \{A \in \Omega \mid \Gamma \vdash_{\mathbf{D}} A\}$.

Definition 6 (**Semantic Consequence**). *For* $\Gamma \subset \mathcal{W}_\Gamma$: $\Gamma \models_{\mathbf{MLA^s}} A$ *(*A is a semantic consequence from Γ*) iff* A *is verified by all simply all right models of* Γ.

The fact that $\mathbf{MLA^s}$ is in standard format warrants that the following theorem holds:[20]

Theorem 1 (**Soundness and Completeness**). $\Gamma \vdash_{\mathbf{MLA^s}} A$ *iff* $\Gamma \models_{\mathbf{MLA^s}} A$.

6 Case Study: The Origin of the Moon

In the first decades after the NASA was founded in 1958, lunar exploration was one of its most prestigious goals. These efforts have led to the Apollo programme that included 6 lunar landing missions between 1969 and 1972.

> There was widespread expectation that the Apollo exploration of the Moon would settle the question of its origin; this had been cited frequently as one of the scientific goals of the Apollo program. (Wood, 1986, p. 18)

As history taught us, this goal was not achieved. Seen in retrospect, one of the most important reasons for the lack of success was

> . . . the concentration on three classical theories of lunar origin: (1) Capture – capture of a planetesimal, formed elsewhere in the solar system, into Earth's orbit; (2) Fission – spontaneous ejection of upper mantle material into a circumterrestrial swarm due to rotational instability, probably during core formation; (3) Coaccretion – formation of the Moon by accretion in a circumterrestrial nebula. (Hartmann, 1986, p. 579)

The main reasons[21] why these hypotheses were considered untenable can be summed up as follows. Firstly, capture (H_1) of a planetesimal can – according to the laws of celestial mechanics – only occur if the original trajectory of this planetesimal is within very limited constraints which include that this protomoon should have originated at about the same (radial) distance from the sun and at about the same time as the Earth. But if they originated at the same time at roughly the same spot in the circumsolar nebula, the moon should have more or less the same chemical composition as the Earth. This is not the case, because the moon hardly contains any iron, one of the heavier elements in the solar system that is abundant in the core of the Earth. Secondly, fission

[20] For an overview of all meta-theoretic properties (and their proofs) of adaptive logics in standard format, we refer to Batens (2007).

[21] As listed in, for instance, the review article of Wood (1986).

(H_2) cannot explain the depletion of volatile elements on the Moon's surface (in comparison with the Earth's surface), neither can it account for the abnormally high angular momentum of the Earth-Moon system (in comparison with other planetary systems in our solar system). Finally, coaccretion (H_3) – that was until then the most supported hypothesis – can neither account for iron depletion nor for the high angular momentum.

Coming to this point several scientists in the mid-seventies were trying to figure out a new hypothesis. Soon, a fourth hypothesis was proposed independently by Hartmann (1975) and Cameron and Ward (1976). We will now focus on the thought process in Cameron and Ward's paper. They started with focussing on the angular momentum of the Earth-Moon system.

> A key constraint on the origin of the Earth-Moon system is the abnormally large value of the specific angular momentum of the system, compared to that of the other planets in the solar system. (p. 120)

Reasoning in terms of elementary dynamics of physical bodies – in which a collision with another body can lead to an increase in angular momentum – they abduced the following hypothesis.

> This spin was presumably imparted by a collision with a major secondary body in the late stages of accumulation of the Earth, with the secondary body adding its mass to the remainder of the proto-Earth. (p. 120)

After determining the characteristics of such a second body – a body of roughly the size of Mars, approaching at 11 km/s and hitting the Earth off center – to account for the specific angular momentum, they could deductively reason further what the consequences would be of such a giant impact. In short, a lot of volatile elements would vaporize upon shock-unloading and a disk of debris would be caught in the gravitational field of the Earth. After a while, the heavier elements (including iron) would sink into the still very fluid young Earth, while the lighter elements that remain in an elliptical trajectory around the Earth would form after a certain amount of time the Moon by accretion. Thus, deriving deductively further consequences of this hypothesis, they concluded that "the Moon should thus be deficient in metallic iron and volatile elements..." (p. 121) and that this hypothesis could at first sight account for all – previously problematic – data.

Before we start to model our case study, it is important to note that we are interested in the process of abduction or the heuristical process of forming explanatory hypotheses, not in confirmation theory or (justificational) inference to the best explanation.[22] So, what we are modelling is the reasoning process

[22]In discerning abduction and IBE I follow the reasoning initiated by Hintikka (1999) and elaborated by Schurz (2008a,b) that the distinction is to be found in their function and context. Abduction is a strategical or heuristical process with minimal justificational value in the context of discovery, while IBE is a justificational process (in the context of justification).

of scientists confronted with counterevidence and looking for a new explanatory hypothesis for the origin of the Moon.[23] This is a different reasoning process than confirmation processes, in which we try to decide whether we have sufficient evidence to support a certain conclusion. This explains the more qualitative nature of arguments in abductive reasoning versus the more quantitative nature of these arguments in justification. We will further use the following notations:

m	"the Moon"
Ox	"x exists in its actual state"
Ax	"x is part of a system with unusually high angular momentum"
Fx	"x has an iron (Fe) core"
Vx	"x is mostly depleted from volatile elements"
Ix	"x is the result of a collision with a secondary body"

Then we can model the relevant background knowledge of Cameron and Ward. The domain of the variables is the set of all natural satellites of our solar system.

1	$\Box Om$	-;PREM	$\emptyset$
2	$\Box\neg Fm$	-;PREM	$\emptyset$
3	$\Box Vm$	-;PREM	$\emptyset$
4	$\Box Am$	-;PREM	$\emptyset$
5	$\Box(\forall x)(Ix \supset Ax)$	-;PREM	$\emptyset$

From these premises they could derive their new hypothesis.

6	$\Diamond Im$	4,5;RC	$\{!Im \triangleright Am\}$

Note that if they would have tried to come up with one of the old hypotheses by considering one of the implications $\Box(\forall x)(H_i x \supset Ox)$ as an extra premise[24] and abducing the corresponding hypothesis $\Diamond H_i m$, these hypotheses would have been defeated. But, as we can see in the following extension of our proof, our new hypothesis $\Diamond Im$ actually predicts all the known (and previously problematic) data about the Moon.

7	$\Box(\forall x)(Ix \supset Ox)$	-;PREM	$\emptyset$
8	$\Box(\forall x)(Ix \supset \neg Fx)$	-;PREM	$\emptyset$
9	$\Box(\forall x)(Ix \supset Vx)$	-;PREM	$\emptyset$
10	$\Diamond Om$	6,7;RU	$\{!Im \triangleright Am\}$
11	$\Diamond\neg Fm$	6,8;RU	$\{!Im \triangleright Am\}$
12	$\Diamond Vm$	6,9;RU	$\{!Im \triangleright Am\}$

[23] New hypotheses can be found both via creative or non-creative abductive processes. As stated in the introduction, this logic does not model creative abductions, which would imply that the conditional used by Cameron and Ward would have been created. Instead, the new hypothesis found by Cameron and Ward is obtained by selecting an existing conditional in their background knowledge ("Collisions have an impact on the angular momentum in systems of physical bodies") and using it to abduce a new hypothesis for the origin of the moon.

[24] Although these three hypotheses are not able to explain the origin of the Moon, some of them are leading hypotheses for other natural satellites in our solar system. See, for instance, Canup and Ward (2002).

Since the new hypothesis is at first sight corroborated by the known data, Cameron and Ward (and other scientists in the field) can now go on and try to justify or prove that this new hypothesis is the actual explanation for the origin of the moon. This also nicely illustrates that it is not possible to sharply distinguish between the context of discovery and the context of justification.[25] There is already in the initial phase of hypothesis formation a justificational aspect present (which we labeled here "corroboration with the known data").[26]

That this modelling of their thought process can be assumed to be accurate follows from the reflection upon their own thought process that Cameron and Ward state in their conclusions.

> We wish to emphasize that this picture follows as a logical consequence of the process needed to provide the angular momentum of the Earth-Moon system. (p. 121)

This conclusion is almost correct, it only omits their abductive move: only if we take the increased angular momentum as a result of a collision, all the other characteristics follow as deductive consequences. Their essential consideration was that collisions are a well-known cause of changes in parameters of dynamic systems.

Together with the independently proposed article by Hartmann (1975),[27] these papers have led to a new successful hypothesis about the origin of the Moon, which has been adopted as the "giant impact hypothesis" (Hartmann, 1986). Increased interest in this problem led to the 1984 conference on the origin of the Moon in Kona. At the conference it became clear that a "major shift of confidence has occurred among lunar scientists" towards the giant impact hypothesis (Wood, 1986, p. 47). At present, this hypothesis is still the most favorable hypothesis amongst lunar scientists (Belbruno and Gott III, 2005, p. 1), although one is still looking for more conclusive evidence by modelling this impact with computer simulations.

7 Conclusions and Open Problems

In this paper, we presented the logic $\mathbf{MLA^s}$ that enables us to model abductive reasoning processes of scientists. Scientists are in general interested in the actual explanation of the puzzling phenomena they investigate. This means that in the case of multiple explanatory hypotheses, scientists will further investigate the different hypotheses one by one. The logic $\mathbf{MLA^s}$ provides this possibility by allowing to derive – in a defeasible way – the different hypotheses. The logic $\mathbf{MLA^s}$ is a decent formal logic in every possible way. Since the logic is formulated in the standard format of adaptive logics, this logic has a proof theory and a semantics that is sound and complete with respect to it.

[25] As discussed and argued in, among others, Aliseda (2006).

[26] For an elaborate discussion on justification in scientific discovery, see Nickles (1980).

[27] This paper mostly explains that such collisions in the initial stadia of our solar system were not as uncommon as thought.

While this logic is apt to model actual abductive processes in science – as our case study points out – several extensions can still enrich this logic. An interesting addition would be that the logic could also handle explananda that are contradictory to our existing background knowledge. Another extension that comes to mind is the ability to handle a structured or layered background knowledge. Finally, there is still a lot of work to be done on the heuristics behind abductive reasoning. Is there a pattern in how scientists find relevant conditionals?

References

A. Aliseda. *Abductive Reasoning.* Springer, Dordrecht, 2006.

D. Batens. The need for adaptive logics in epistemology. In D. Gabbay, S. Rahman, J. Symons, and J.P. Van Bendegem, editors, *Logic, Epistemology and the Unity of Science*, pages 459–485. Kluwer Academic Publishers, Dordrecht, 2004.

D. Batens. A universal logic approach to adaptive logics. *Logica Universalis*, 1: 221–242, 2007.

D. Batens. *Adaptive Logics and Dynamic Proofs. A Study in the Dynamics of Reasoning.* Forthcoming, 2011.

E. Belbruno and J. R. Gott III. Where did the moon come from? *The Astronomical Journal*, 129(3):1724–1745, 2005.

A.G.W. Cameron and W.R. Ward. The origin of the moon. In *Abstracts of the Lunar and Planetary Science Conference*, pages 120–122, Houston, 1976. Lunar Science Institute.

R. M. Canup and W. R. Ward. Formation of the Galilean satellites: Conditions of accretion. *The Astronomical Journal*, 124(6):3404–3423, 2002.

W. K. Hartmann. Moon origin: The impact-trigger hypothesis. In W.K. Hartmann, Phillips R.J., and G.J. Taylor, editors, *Origin of the moon; Proceedings of the Conference, Kona, HI, October 13-16, 1984*, pages 17–56, Houston, 1986. Lunar and Planetary Institute.

W.K. Hartmann. Satellite-sized planetesimals and lunar origin. *Icarus*, 24: 504–515, 1975.

J. J. Hintikka. What is abduction? The fundamental problem of contemporary epistemology. In *Inquiry as Inquiry: a Logic of Scientific Discovery*, volume 5 of *Jaakko Hintikka Selected Papers*, pages 91–113. Kluwer Academic Publishers, Dordrecht, 1999.

H. Lycke. The adaptive logics approach to abduction. In E. Weber, T. Libert, P. Marage, and G. Vanpaemel, editors, *Logic, Philosophy and History of Science in Belgium. Proceedings of the Young Researchers Days 2008*, pages 35–41, Brussels, 2009. KVAB.

J. Meheus. Adaptive Logics for Abduction and the Explication of Explanation-Seeking Processes. In Olga Pombo and Alexander Gerne, editors, *Abduction and the Process of Scientific Discovery*, pages 97–119. Centro de Filosofia das Cincias da U. de Lisboa, Lisboa, 2007.

J. Meheus. A formal logic for the abductions of singular hypotheses. *Forthcoming*, 2010. Online manuscript on http://logica.ugent.be/centrum/writings/pubs.php.

J. Meheus and D. Batens. A formal logic for abductive reasoning. *Logic Journal of the IGPL*, 14:221–236, 2006.

T. Nickles. *Scientific Discovery, Logic, and Rationality*, chapter Scientific Discovery and the Future of Philosophy of Science. D. Reidel Publishing Company, Dordrecht, 1980.

G. Schurz. Patterns of Abduction. *Synthese*, 164:201–234, 2008a.

G. Schurz. Common Cause Abduction and the Formation of Theoretical Concepts in Science. In C. Dégremont, L. Keiff, and H. Rückert, editors, *Dialogues, Logics, and Other Strange Things. Essays in Honour of Shahid Rahman*, pages 337–364. College Publications, London, 2008b.

J. A. Wood. Moon over mauna loa: a review of hypotheses of formation of earth's moon. In W.K. Hartmann, Phillips R.J., and G.J. Taylor, editors, *Origin of the moon; Proceedings of the Conference, Kona, HI, October 13-16, 1984*, pages 579–608, Houston, 1986. Lunar and Planetary Institute.