

A Formal Logic for the Abduction of Singular Hypotheses

Joke Meheus

Centre for Logic and Philosophy of Science
University of Ghent, Belgium

`Joke.Meheus@UGent.be`

This is the final version of the paper as published in:

Dennis Dieks, Wenceslao J. Gonzalez, Stephan Hartmann, Thomas Uebel and Marcel Weber (eds.), *Explanation, Prediction, and Confirmation*, Springer, 2011, pp. 93-108.

DOI 10.1007/978-94-007-1180-8_6

A Formal Logic for the Abduction of Singular Hypotheses*

Joke Meheus

Centre for Logic and Philosophy of Science
University of Ghent, Belgium

Joke.Meheus@UGent.be

April 2, 2010

1 Introduction

The aim of this paper is to present a new logic, called \mathbf{LA}_s^r , for the abduction of singular hypotheses. The intended application context of the logic concerns problem solving situations in which one tries to ‘abduce’ an explanatory hypothesis for an *explanandum* on the basis of a background theory possibly together with some relevant empirical findings. The intended set of premises will thus typically consist of one or more *explananda* and some general statements that, with respect to those particular *explananda*, are considered as relevant. It may also contain zero or more empirical findings that are relevant (for instance, because they contradict one of the *explanantia*).

The logic presented here is a variant of the logic \mathbf{LA}^r .¹ Both \mathbf{LA}^r and \mathbf{LA}_s^r are adaptive logics. Adaptive logics are a family of non-standard logics that are meant to study defeasible reasoning processes in a formally exact way. As will become clear below, the adaptive logics framework is especially suited to formulate formal logics for abduction.²

The logics \mathbf{LA}^r and \mathbf{LA}_s^r lead to the same consequence set when applied to the same set of premises. The difference, however, is that \mathbf{LA}_s^r , unlike \mathbf{LA}^r , is in the so-called “standard format” of adaptive logics.

Formulating an adaptive logic in standard format has many advantages. One of them is that both the semantics and the proof theory can be formulated in a

*Research for this paper was supported by subventions from Ghent University and from the Research Foundation – Flanders (FWO - Vlaanderen). The author is indebted to the anonymous referees for their helpful suggestions.

¹The logic \mathbf{LA}^r was first presented in Joke Meheus and Diderik Batens, “A Formal Logic for Abductive Reasoning”, in *Logic Journal of the IGPL* 14, 2006, pp. 221-236.

²The first logic in this family was designed around 1980 by Diderik Batens 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, such as abduction and inductive generalization) and a whole variety of adaptive logics was designed—for an excellent introduction to adaptive logics see Diderik Batens, *Adaptive Logics and Dynamic Proofs. Mastering the Dynamics of Reasoning, with Special Attention to Handling Inconsistency*, forthcoming.

completely generic way. A second, and more important advantage, is that the standard format warrants that all central meta-theoretic properties (soundness, completeness, proof invariance, fixed point property, ...) hold. As soon as it can be shown that the adaptive logic satisfies the requirements of the standard format, all these properties can be proven in a generic way (that is, without referring to the properties of the specific adaptive logic). The standard format will be discussed in Section 5.

A final advantage of \mathbf{LA}_s^r is that it is much more transparent and much more user-friendly than \mathbf{LA}^r . One of the reasons for this is that, in \mathbf{LA}_s^r , it is not required to determine beforehand the set of possible *explanantia* and the set of possible *explananda*, which is a drawback of \mathbf{LA}^r .³

Both \mathbf{LA}^r and \mathbf{LA}_s^r are based on Classical Logic—henceforth **CL**. So, all references to causality, laws of nature, and similar non-extensional concepts will be out of the picture. I do not doubt that more interesting results may be obtained from intensional logics. However, as a starting point, I want to keep the discussion as simple and transparent as possible.

2 The problem

Many formal logicians display disinterest or even suspicion with respect to abduction. There seem to be several reasons for this. The first is that abductive steps are of the form

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

a fallacy known as Affirming the Consequent (given Universal Instantiation). The second is that many examples of purportedly sound abductions seem to rely on a hidden non-formal reasoning: the only sensible formal rule behind them seems to lead inevitably to a set of unsound and even inconsistent conclusions. For instance, given the *explananda* Qa and Ra and the generalizations $(\forall x)(Px \supset Qx)$ and $(\forall x)(\neg Px \supset Rx)$, (\dagger) enables one to infer both Pa and $\neg Pa$. A third reason seems to be that abductive inferences are ‘contextually validated’. For instance, although it makes sense to infer Pa on the basis of $\{(\forall x)(Px \supset Qx), Qa, Qb \wedge \neg Pb\}$, it does not make sense to infer Pb from that same set of premises.

Still, mainly from the side of computer science, several attempts have been made at formulating a logic-based approach to abduction.⁴ What these approaches have in common is that (\dagger) 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 that⁵

- (1) $T \cup \{A\} \vdash B$.
- (2) $T \not\vdash B$.
- (3) $T \not\vdash \neg A$.

³That both sets have to be determined beforehand is one of the reasons that \mathbf{LA}^r is not in standard format—see Section 5 for the requirements on the standard format.

⁴For an excellent introduction to logic-based approaches to abduction, see Atocha Aliseda, *Abductive Reasoning. Logical Investigations into Discovery and Explanation*. Dordrecht: Springer 2006.

⁵I use the notion “theory” here in a broad sense. It contains the background theory, but may also contain empirical findings that are relevant for the explanation problem.

- (4) $B \not\vdash A$.
- (5) A is ‘minimal’.

The first of these requirements needs little 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 .⁶ (4) is needed to rule out (partial) self-explanations. 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 an *explanandum*, one is interested in explanations that are as *parsimonious* as possible. Hence, in view of $(\forall \alpha)(A(\alpha) \supset B(\alpha)) \vdash_{\mathbf{CL}} (\forall \alpha)((A(\alpha) \wedge C(\alpha)) \supset B(\alpha))$, for an arbitrary C , (the so-called property of strengthening the antecedent) one needs to prevent that $A(\beta) \wedge C(\beta)$ can be abduced, whenever $A(\beta)$ can. This can be realized by requiring that the explanatory hypothesis is ‘minimal’.

The 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 and that also satisfies the requirements (1)–(4). This is the option that is followed for the logic \mathbf{LA}_s^r .

The approach in the present paper is very different from the existing logic-based approaches to abduction. As mentioned above, the logic \mathbf{LA}_s^r is an adaptive logic. Defining a logic for abduction within the adaptive logic framework has several advantages. A first one is that the resulting logics (unlike the systems available today) have 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, abductive reasoning is not only undecidable, there even is no positive test for it (see also Section 3).⁷

Another advantage of the proposed logics is that they are much closer to natural reasoning than the existing systems. As we mentioned before, 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 proposed in this paper treats abduction as a form of ‘forward reasoning’: it is an ampliative system that directly validates certain inferences of the form (\dagger).

The third advantage is related to this: deductive and abductive steps are nicely integrated into a single system. As a consequence, the logics not only enable one to generate explanatory hypotheses, but also to infer predictions on the basis of explanatory hypotheses and the background theory. This is

⁶A formula A is said to be compatible with a set of premises Γ iff $\Gamma \not\vdash \neg A$.

⁷A property for objects of a given kind is *decidable* iff there is a mechanical procedure that leads to the answer YES if the property holds and to the answer NO if the property does not hold. There is a positive test for objects of a given kind iff there is a mechanical procedure that leads to the answer YES if the property holds. If the property does not hold the procedure may lead to the answer NO, but may continue forever. So, even if A follows abductively from a theory T and an *explanandum* B , there need not exist any finite construction that establishes this.

highly important from the point of view of applications. In all interesting cases, explanatory hypotheses are typically used to derive predictions which, in turn, may lead to a revision of the original hypotheses.

A final advantage is that, unlike what is the case for the existing logic-based approaches, \mathbf{LA}_s^r does not require that one determines beforehand the set of potential explanations. This makes it a much more flexible tool, especially for cases that involve ill-structured problems, where it may not be clear at all what would count as an explanation.

3 Main Characteristics of Abductive Reasoning

In order to present the logic \mathbf{LA}_s^r , I shall first discuss in some more detail the main characteristics of abductive reasoning processes.

As mentioned in the previous section, abductive inferences should lead to conclusions that are compatible with the premises. It is important to note, however, that different *explanantia* are not necessarily *jointly* compatible with the premises. None of the requirements (1)–(5) excludes that different explanations are incompatible with each other. Evidently, this raises the question how one can avoid, in a classical framework, that the generation of contradicting explanations leads to triviality. We shall see below that this problem is handled in a very natural way by the logic \mathbf{LA}_s^r .

A second characteristic is that abduction is a *non-monotonic* form of reasoning: conclusions that follow abductively from a theory T may be withdrawn when T is extended to $T \cup T'$. This characteristic is related to the fact that some of the requirements for abductive inferences are *negative*—see (2)–(4) in the previous section.

To see the relation between the negative requirements and the non-monotonic character of abductive inferences more clearly, consider the following simple example:

- (6) John has a fever and small red spots on his face and body.
- (7) Everybody who has rubeola (the measles) has a fever and small red spots on the face and body.
- (8) Everybody who has rubella (the German measles) has a fever and small red spots on the face and body.
- (9) Nobody has rubeola more than once.
- (10) Nobody has rubella more than once.

Suppose that (6) is the *explanandum* B , and that (7)–(10) constitute the background theory. From (7), we may derive by **CL**:

- (11) If John has rubeola, then he has a fever and small red spots on his face and body.

Hence, as requirements (1)–(5) are evidently fulfilled for (6), (7)–(10), and the antecedent of (9), we may abduce

- (12) John has rubeola.

Similarly, we may add on the basis of (8)

- (13) If John has rubella, then he has a fever and small red spots on his face and body.

This in turn allows us to abduce

- (14) John has rubella.

So, we derived two different *explanantia* for the same *explanandum*. However, if we interpret a ‘*minimal* explanation’ as the logically weakest one, then neither of the two *explanantia* satisfies this criterion, but their disjunction does. It is easy to see how this disjunction may be derived in the proof. From (7) and (8) follows

- (15) Everybody who has rubeola or rubella has a fever and small red spots on the face and body.

From this follows by **CL**

- (16) If John has rubeola or rubella, then he has a fever and small red spots on his face and body.

But then, in view of (6) and (15) we may abduce

- (17) John has rubeola or he has rubella.

In view of the disjunction of possible explanations on line (17), the explanations on lines (12) and (14) are no longer minimal. In \mathbf{LA}_s^r , from line (17) on, lines (12) and (14) are *marked*, indicating that the sentences that occur on them are no longer considered as derived in the proof (see Sections 4 and 5 to see how this marking is technically realized).

As was mentioned in the introduction, ampliative reasoning processes display an interesting interaction between abductive steps and deductive steps.

Suppose that we continue our previous proof with a relevant finding that we just found out about John:

- (18) John had rubeola last year.

But then, from (9) and (18) we may derive deductively

- (19) John does not have rubeola.

And hence, in view of (17) and (19), we can rule out one of the potential explanations, leaving us in this case with one potential explanation:

- (17) John has rubella.

The final characteristic is that abductive reasoning processes, at the predicative level, do not have a *positive test*. This is related to the fact that first-order predicate logic is undecidable—if a conclusion A does *not* follow from a set of premises Γ , we may not be able to establish this (see also footnote 7). Hence, as abductive inferences are partly defined in terms of negative requirements, it immediately follows that, for undecidable fragments of first-order logic, abductive inferences lack a positive test. Suppose, for instance, that for a theory T ,

an *explanandum* B , and a sentence A , (1) is satisfied. In that case, it seems reasonable to conclude that A follows abductively from T , *provided* that (2)–(5) are satisfied. However, there is no algorithm to establish. Hence, if one is unable to establish that (2)–(5) are satisfied, there is no absolute warrant that the abductive step is not erroneous.

This brings us to the general idea behind our adaptive logic. In \mathbf{LA}_s^r it is allowed that, as soon as (1) is satisfied, an abductive step is made, but *on the condition* that (2)–(5) are satisfied. In other words, for any possible application of abduction, it is assumed, unless and until proven otherwise, that the conditions (2)–(5) are satisfied. How this is realized technically will become clear below. We shall end this section with a motivation for this general idea.

There are different ways to deal with the lack of a positive test. The most common one 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 unacceptable restriction: nearly all interesting theories are such that their set of theorems is undecidable.

The alternative 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 a theory T , an *explanandum* B , and a sentence A , one establishes that (1) is satisfied, and one did not establish that one of (2)–(5) is violated. In line with the general idea from two paragraphs ago, we may consider A as an abductive consequence of 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 establishes 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 (by adding new premises) or by an internal one (by gaining better insight in the premises) does not seem essential. The third, and most important argument is that, even for decidable fragments of first-order logic, it is often unrealistic to require absolute warrants. Even if a decision method is available, reasoners may lack the resources to perform an extensive search, and hence, may be forced to act on their present best insights.

The logic \mathbf{LA}_s^r follows this second option. This has the advantage that, even for undecidable fragments of first-order logic, 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.⁸

⁸At this point, one may interpret “an abductive conclusion A is *finally derivable* from a theory T ” as “ A satisfies the requirements (1)–(5)” —see Section 5 for a precise definition of

4 Informal Presentation of the Logic \mathbf{LA}_s^r

The general idea behind \mathbf{LA}_s^r is extremely simple: given a set of premises Γ , it is allowed that any valid inference rule of \mathbf{CL} is applied (\mathbf{LA}_s^r is an adaptive extension of \mathbf{CL}), and moreover that (\dagger) 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 \mathbf{CL} -derivable from Γ ”. So, whenever it is \mathbf{CL} -derivable from Γ that, for some general rule $(\forall\alpha)(A(\alpha) \supset B(\alpha))$ and some *explanandum* $B(\beta)$, (\dagger) cannot be applied consistently (because, $\neg A(\beta)$ is \mathbf{CL} -derivable from Γ), the application of (\dagger) is overruled. In view of what we have seen in Section 2 (requirement (3)), this is exactly what we want.

A formula of the form $(\forall\alpha)(A(\alpha) \supset B(\alpha)) \wedge (B(\beta) \wedge \neg A(\beta))$ will henceforth be called an *abnormality*. It is a formula that *blocks* a desired inference (in this case an abductive step). In line with what is common for adaptive logics, we shall say that \mathbf{LA}_s^r interprets sets of premises “as normally as possible”. In this case this means that abductive steps are “applied as much as possible”.⁹ To keep the proofs below within limits, I shall use $(\forall\alpha)(A(\alpha) \supset B(\alpha))(\alpha/\beta)$ as an abbreviation for formulas of the form $(\forall\alpha)(A(\alpha) \supset B(\alpha)) \wedge (B(\beta) \wedge \neg A(\beta))$.

In order to rule out (partial) self-explanations, we only need one refinement. Given a formula $(\forall\alpha)(A(\alpha) \supset B(\alpha))(\alpha/\beta)$, we shall require that $A(\beta)$ does not share any predicate with $B(\beta)$. So, none of the following formulas counts as an abnormality:

$$\begin{aligned} & (\forall x)(Px \supset Px) \wedge (Pa \wedge \neg Pa) \\ & (\forall x)((Px \wedge Qx) \supset Px) \wedge ((Pa \wedge Qa) \wedge \neg(Pa \wedge Qa)) \\ & (\forall x)((Px \vee Qx) \supset (Px \vee Rx)) \wedge ((Pa \vee Ra) \wedge \neg(Pa \vee Ra)) \\ & (\forall x)((Px \wedge Qx) \supset (Px \wedge Rx)) \wedge ((Pa \wedge Ra) \wedge \neg(Pa \wedge Ra)) \end{aligned}$$

From now on, we shall only use the term “abnormality” for formulas of the form $(\forall\alpha)(A(\alpha) \supset B(\alpha))(\alpha/\beta)$ that satisfy this additional requirement.

To realize the contextual validation of (\dagger) , we shall make a distinction in the proof theory between unconditional rules and conditional rules. The unconditional rules are all those that are validated by \mathbf{CL} . The conditional rules are applications of (\dagger) . In order to avoid referring to a large set of specific rules, and as is common for adaptive logics, I shall use two *generic* rules (see below for illustrations and Section 5 for the precise definitions of the rules). The generic rule for the unconditional rules is called RU, and the one for the conditional rules RC.

Here is a very simple example of a proof in \mathbf{LA}_s^r . The fifth element in the proof is the so-called condition—its function will become clear immediately.

1	$(\forall x)(Px \supset Qx)$	PREM	\emptyset
2	$(\forall x)(\neg Px \supset Rx)$	PREM	\emptyset
3	$(\forall x)(Sx \supset Tx)$	PREM	\emptyset
4	$(\forall x)(Sx \supset Vx)$	PREM	\emptyset
5	$(\forall x)(Wx \supset \neg Sx)$	PREM	\emptyset
6	$Qa \wedge Ra$	PREM	\emptyset

the notion of *final derivability*.

⁹The term “abnormality” is a technical term that is common in the literature on adaptive logics and that is application dependent. It does not refer to any standard of normality.

7	$Ta \wedge Tb$	PREM	\emptyset
8	Wa	PREM	\emptyset

The formulas on lines 1–5 correspond to the background theory; those on lines 6–8 to *explananda*. In view of the formulas on lines 1 and 6 we may derive

9	Pa	1,6; RC	$\{(\forall x)(Px \supset Qx)(x/a)\}$
---	------	---------	---------------------------------------

Note that the formula $Pa \vee (\forall x)(Px \supset Qx)(x/a)$ is **CL**-derivable from the premises.¹⁰ Hence, one may read the latter formula as: Pa is derivable from the premises *provided* $(\forall x)(Px \supset Qx)(x/a)$ is false. This is why it is allowed in the adaptive proof that Pa is derived on the condition $\{(\forall x)(Px \supset Qx)(x/a)\}$. The underlying idea is that abnormalities are supposed to be false, unless and until proven otherwise. If at some stage the condition of line 9 is no longer satisfied, the line is marked and the formula that occurs on that line is no longer considered as derived. The addition of “RC” in the justification refers to the fact that this line was added by the generic conditional rule.

Note also that potential explanations will always be introduced on a non-empty condition. This provides an easy way to distinguish, in the proof, between **CL**-consequences of the theory and potential explanations.

In view of the formulas on lines 2 and 6, we may continue the proof as follows

10	$\neg Pa$	2,6; RC	$\{(\forall x)(\neg Px \supset Rx)(x/a)\}$
----	-----------	---------	--

This brings us immediately to a slight complication, that is best dealt with first. From the premises, neither $(\forall x)(Px \supset Qx)(x/a)$ nor $(\forall x)(\neg Px \supset Rx)(x/a)$ is **CL**-derivable. However, their disjunction is:

11	$(\forall x)(Px \supset Qx)(x/a) \vee (\forall x)(\neg Px \supset Rx)(x/a)$	1,2,6; RU	\emptyset
----	---	-----------	-------------

The information we obtain from the formula on line 11 is that, in view of the premises, one of the abnormalities $(\forall x)(Px \supset Qx)(x/a)$ and $(\forall x)(\neg Px \supset Rx)(x/a)$ is true. However, at this stage of the proof, we do not know which one is. Another way to put this is that, in view of the premises, there is no reason to choose one abnormality above the other, and hence, also no reason to choose the explanatory hypothesis Pa above the hypothesis $\neg Pa$.

It is in view of cases like this that a *strategy* is needed to disambiguate the phrase “to interpret the premises as normally as possible”. The strategy that is followed in the case of \mathbf{LA}_s^t is the so-called Reliability Strategy. This requires some definitions and conventions on notation.

Where Δ is a finite set of abnormalities, the disjunction $\bigvee(\Delta)$ will be called a *Dab*-formula and will be written as $Dab(\Delta)$. A *Dab*-formula $Dab(\Delta)$ will be called a *minimal Dab*-formula at stage s of a proof, if, at that stage of the proof, no $Dab(\Delta')$ is derived, such that $\Delta' \subset \Delta$. As will become clear below, the marking of lines at a stage s of a proof will proceed in terms of the minimal *Dab*-formulas that are derived at that stage of the proof.

What the Reliability Strategy comes to is that, whenever a *minimal Dab*-formula is derived in the proof at a certain stage (as is the case on line 11 of

¹⁰I say that a formula A is *derived* in a proof from Γ if A occurs on some unmarked line in the proof. I say that a formula A is *derivable* from Γ if there exists a proof from Γ such that A occurs on an unmarked line in it. In Section 5, we shall see the precise definitions for *derived at a stage*, *finally derived at a stage* and *finally derivable*.

our proof), then all disjuncts that occur in that *Dab*-formula are considered as *unreliable*, and hence, all lines that rely on the normal behaviour of any of these disjuncts are marked. Or, put differently, a line is marked at a stage s if its condition contains an abnormality that is a disjunct of a minimal *Dab*-formula at that stage. In our proof, this means that both lines 9 and 10 are marked from stage 11 on and the formulas that occur on these lines are no longer considered as derived in the proof. This is precisely what we want, because we want our set of *explanantia* to be jointly compatible with the premises. This is how the proof looks at stage 11:

1	$(\forall x)(Px \supset Qx)$	PREM	\emptyset
2	$(\forall x)(\neg Px \supset Rx)$	PREM	\emptyset
3	$(\forall x)(Sx \supset Tx)$	PREM	\emptyset
4	$(\forall x)(Sx \supset Vx)$	PREM	\emptyset
5	$(\forall x)(Wx \supset \neg Sx)$	PREM	\emptyset
6	$Qa \wedge Ra$	PREM	\emptyset
7	$Ta \wedge Tb$	PREM	\emptyset
8	Wa	PREM	\emptyset
9	Pa	1,6; RC	$\{(\forall x)(Px \supset Qx)(x/a)\} \checkmark^{11}$
10	$\neg Pa$	2,6; RC	$\{(\forall x)(\neg Px \supset Rx)(x/a)\} \checkmark^{11}$
11	$(\forall x)(Px \supset Qx)(x/a) \vee$ $(\forall x)(\neg Px \supset Rx)(x/a)$	1,2,6; RU	\emptyset

The following extension of the proof illustrates the way in which abductive steps are contextually validated in \mathbf{LA}_s^r . Suppose that we continue the proof as follows

12	Sa	3,7; RC	$\{(\forall x)(Sx \supset Tx)(x/a)\}$
13	Sb	3,7; RC	$\{(\forall x)(Sx \supset Tx)(x/b)\}$

However, the following line may be added by the rule RU:

14	$\neg Sa$	5,8; RU	\emptyset
----	-----------	---------	-------------

At this point it becomes clear that the following abnormality is **CL**-derivable

15	$(\forall x)(Sx \supset Tx)(x/a)$	3,7,14; RU	\emptyset
----	-----------------------------------	------------	-------------

From this point on, the condition of line 12 overlaps with the minimal *Dab*-formula on line 15, and hence, this line is marked:

...			
12	Sa	3,7; RC	$\{(\forall x)(Sx \supset Tx)(x/a)\} \checkmark^{15}$
13	Sb	3,7; RC	$\{(\forall x)(Sx \supset Tx)(x/b)\}$
14	$\neg Sa$	4,5; RU	\emptyset
15	$(\forall x)(Sx \supset Tx)(x/a)$	3,7,14; RU	\emptyset

Despite the fact that Sa cannot be abduced on the basis of lines 3 and 7, Sb can be abduced on the basis of these same lines. For this simple example, it is easy to check that line 15 will not be marked in any extension of the proof (Sb is *finally derived* in this proof—see Section 5 for the precise definition of final derivability).

This is also a good point to illustrate how abductive steps and deductive steps are nicely integrated in \mathbf{LA}_s^r . In view of lines 4 and 13, the following line may be added:

16 Vb 4,13; RU $\{(\forall x)(Sx \supset Tx)(x/b)\}$

The formula on line 16 is a prediction that is **CL**-derivable from the premises and the explanatory hypothesis on line 13. The line is added by the generic unconditional rule RU. Note that when the unconditional rule is applied, no new formulas are added to the condition, but any formula that occurs in a non-empty condition is ‘carried’ over to the conclusion of the application. The reason for this is easy to understand. If, at some point, line 13 has to be marked (because its condition is no longer satisfied), then evidently any line that depends on it, should also be marked.

I shall now show that the logic \mathbf{LA}_s^r has some further desired properties. The first is that no logically contingent formula is finally derivable as an explanation for a tautology. Suppose that we would start a proof by introducing two **CL**-theorems:

1 $Pa \vee \neg Pa$ RU \emptyset
 2 $(\forall x)(Qx \supset (Px \vee \neg Px))$ RU \emptyset

We may now continue the proof as follows

3 Qa RU $\{(\forall x)(Qx \supset (Px \vee \neg Px))(x/a)\}$

However, this is harmless. As soon as we add the following line:

4 $(\forall x)(Qx \supset (Px \vee \neg Px))(x/a) \vee (\forall x)(\neg Qx \supset (Px \vee \neg Px))(x/a)$ RU \emptyset

line 3 is marked.

There is also no risk that contradictions would be (finally) abducible for logically contingent formulas. Suppose that our *explanandum* is Qa and that we introduce the **CL**-theorem $(\forall x)((Px \wedge \neg Px) \supset Qx)$ in our proof:

1 Qa PREM \emptyset
 2 $(\forall x)((Px \wedge \neg Px) \supset Qx)$ RU \emptyset

We may now abduce $Pa \wedge \neg Pa$ as an explanation for Qa , but this line is marked as soon as we unconditionally derive $(\forall x)((Px \wedge \neg Px) \supset Qx)(x/a)$ from Qa in the proof:

3 $Pa \wedge \neg Pa$ 1,2; RC $\{(\forall x)((Px \wedge \neg Px) \supset Qx)(x/a)\}$ \checkmark^4
 4 $(\forall x)((Px \wedge \neg Px) \supset Qx)(x/a)$ 1; RU \emptyset

The derivation of (partial) self-explanations is ruled out by the restriction that we have imposed upon our abnormalities. Consider the following proof, where Pa is our *explanandum*

1 Pa PREM \emptyset
 2 $(\forall x)((Px \vee Qx) \supset Px)$ PREM \emptyset

Evidently, we may extend the proof by

3 $(Pa \vee Qa) \vee ((\forall x)((Px \vee Qx) \supset Px)(x/a))$ 1; RU \emptyset

However, in this case, the derivation of line 3 will not allow us to derive $Pa \vee Qa$ on the condition $(\forall x)((Px \vee Qx) \supset Px)(x/a)$. The reason is that the latter formula is not an abnormality and that the condition may only contain abnormalities (see the next section for the precise definitions). (The formula $Pa \vee Qa$ could evidently be derived on the empty condition by the rule RU from the formula on line 1, but this is obviously harmless. As was mentioned above, only formulas that are derived on non-empty conditions are considered as possible explanations for some *explanandum*.)

There is one last complication we have to address. What if more than one *explanans* is abducible for the same *explanandum*? Consider, for instance, the following proof:

1	$(\forall x)(Px \supset Rx)$	PREM	\emptyset
2	$(\forall x)(Qx \supset Rx)$	PREM	\emptyset
3	Ra	PREM	\emptyset
4	Pa	1,3; RC	$\{(\forall x)(Px \supset Rx)(x/a)\}$
5	Qa	2,3; RC	$\{(\forall x)(Qx \supset Rx)(x/a)\}$

From these premises, the following minimal *Dab*-formulas

6	$(\forall x)(Px \supset Rx)(x/a) \vee (\forall x)((Qx \wedge \neg Px) \supset Rx)(x/a)$	1-3; RU	\emptyset
7	$(\forall x)(Qx \supset Rx)(x/a) \vee (\forall x)((Px \wedge \neg Qx) \supset Rx)(x/a)$	1-3; RU	\emptyset

are **CL**-derivable. This means that, by the Reliability Strategy, both lines 4 and 5 are marked. Hence, as lines 4 and 5 will remain marked in any extension of the proof, neither Pa nor Qa is finally derivable from these premises.¹¹ However, their disjunction is, as is shown by the following extension:

8	$(\forall x)((Px \vee Qx) \supset Rx)$	1,2; RU	\emptyset
9	$Pa \vee Qa$	3,8; RC	$\{(\forall x)((Px \vee Qx) \supset Rx)(x/a)\}$

As no minimal *Dab*-formula is derivable from the premises that has $(\forall x)((Px \vee Qx) \supset Rx)(x/a)$ as one of its disjuncts, the formula on line 9 is finally derived in this proof.

Before ending this section, I want to illustrate that \mathbf{LA}_s^r leads to the most parsimonious explanatory hypotheses. Consider the following example:

1	$(\forall x)(Px \supset Rx)$	PREM	\emptyset
2	$(\forall x)((Px \wedge Qx) \supset Rx)$	PREM	\emptyset
3	Ra	PREM	\emptyset

At this stage both Pa and $Pa \wedge Qa$ may be abduced:

4	Pa	1,3; RU	$(\forall x)(Px \supset Rx)(x/a)$
5	$Pa \wedge Qa$	2,3; RU	$(\forall x)((Px \wedge Qx) \supset Rx)(x/a)$

¹¹The only way in which lines 4 and 5 would be unmarked in some extension is when the *Dab*-formulas on lines 6 and 7 are no longer minimal. However, for this simple example, it is easily seen that neither of the disjuncts of the formula on line 6 is **CL**-derivable from the premises and analogously for the formula on line 7.

However, line 5 is marked as soon as the following minimal *Dab*-formula is derived

$$6 \quad (\forall x)((Px \wedge Qx) \supset Rx)(x/a) \vee (\forall x)((Px \wedge \neg Qx) \supset Rx)(x/a) \quad 1; \text{RU} \quad \emptyset$$

I end this section with a possible objection. Some readers might object that, given a generalization of the form $(\forall \alpha)(A(\alpha) \supset B(\alpha))$, the logic \mathbf{LA}_s^r cannot distinguish between cases in which nearly all B are also A and cases in which very few B are A . In the former case, the argument might continue, an application of abduction for a particular *explanandum* seems sound (it leads to a sufficiently plausible hypothesis), whereas it does not seem sound in the latter case.

My reaction to this objection is twofold. The first is that, in cases where there are two or more possible explanations for the same *explanandum*, only their *disjunction* is \mathbf{LA}_s^r -derivable. So, the plausibility of the (disjunctions of) possible explanations that are \mathbf{LA}_s^r -derivable are at least as plausible as their most plausible disjunct. The second reaction is that the objection seems to rely on a confusion. The aim of \mathbf{LA}_s^r is not to select those *explanantia* that are most plausible. Its aim is to derive, for each *explanandum*, a unique disjunction of possible explanations that is complete (also the most unlikely ones occur in it). For certain application contexts, where it is crucial not to overlook any of the possibilities, this procedure makes good sense.

5 The Logic \mathbf{LA}_s^r

In this section, I present the logic \mathbf{LA}_s^r in a formally precise way. As any other adaptive logic in standard format, the logic \mathbf{LA}_s^r is characterized by a triple: a lower limit logic (a reflexive, transitive, monotonic, uniform, and compact logic for which there is a positive test), a set of abnormalities Ω (characterized by a, possibly restricted, logical form) and a strategy. The lower limit logic is the logic that determines the stable part of the adaptive logic, and that also determines the unconditional rule. In the case of \mathbf{LA}_s^r , the lower limit logic is \mathbf{CL} and the strategy is Reliability.

Relying on the previously introduced abbreviation, the set of abnormalities Ω is defined by

Definition 1 $\Omega = \{(\forall \alpha)((A(\alpha) \supset B(\alpha))(\alpha/\beta) \mid \text{no predicate that occurs in } B(\beta) \text{ occurs in } A(\beta))\}$

In order to define the semantics, we need some further definitions. We first define the abnormal part of a \mathbf{CL} -model:

Definition 2 $Ab(M) = \{A \in \Omega \mid M \Vdash A\}$

We shall say that a *Dab*-formula $Dab(\Delta)$ is a *Dab*-consequence of Γ if it is \mathbf{CL} -derivable from Γ and that it is a *minimal Dab*-consequence of Γ if there is no $\Delta' \subset \Delta$ such that $Dab(\Delta')$ is also a *Dab*-consequence of Γ . The set of formulas that are *unreliable* with respect to Γ , denoted by $U(\Gamma)$, is defined by

Definition 3 *Where $Dab(\Delta_1)$, $Dab(\Delta_2)$, \dots are the minimal *Dab*-consequences of Γ , $U(\Gamma) = \Delta_1 \cup \Delta_2 \cup \dots$ is the set of formulas that are unreliable with respect to Γ .*

In view of these definitions, the semantic consequence relation of \mathbf{LA}_s^r is given by:

Definition 4 A \mathbf{CL} -model M of Γ is reliable iff $Ab(M) \subseteq U(\Gamma)$.

Definition 5 $\Gamma \vDash_{\mathbf{LA}_s^r} A$ iff A is verified by all reliable models of Γ .

As is common for all adaptive logics in standard format, the proof theory of \mathbf{LA}_s^r is characterized by three generic inference rules and a marking definition. The inference rules only refer to the lower limit logic, in our case \mathbf{CL} . Where Γ is the set of premises, the inference rules are given by

$$\begin{array}{ll}
\text{PREM} & \text{If } A \in \Gamma: \quad \frac{\dots \quad \dots}{A \quad \emptyset} \\
\\
\text{RU} & \text{If } A_1, \dots, A_n \vdash_{\mathbf{CL}} B: \quad \frac{\begin{array}{l} A_1 \quad \Delta_1 \\ \dots \quad \dots \\ A_n \quad \Delta_n \\ \hline B \quad \Delta_1 \cup \dots \cup \Delta_n \end{array}}{} \\
\\
\text{RC} & \text{If } A_1, \dots, A_n \vdash_{\mathbf{CL}} B \vee Dab(\Theta) \quad \frac{\begin{array}{l} A_1 \quad \Delta_1 \\ \dots \quad \dots \\ A_n \quad \Delta_n \\ \hline B \quad \Delta_1 \cup \dots \cup \Delta_n \cup \Theta \end{array}}{}
\end{array}$$

The premise rule PREM simply states that, at any line of a proof, a premise may be introduced on the empty condition. What the unconditional rule RU comes to is that, whenever $A_1, \dots, A_n \vdash_{\mathbf{CL}} B$ and the A_1, \dots, A_n occur in the proof on the conditions $\Delta_1, \dots, \Delta_n$, then B may be added to the proof on the condition $\Delta_1 \cup \dots \cup \Delta_n$. The conditional rule RC is analogous, except that here a new condition is introduced.

The marking definition proceeds in terms of the *minimal Dab-formulas* derived at a stage of the proof:

Definition 6 $Dab(\Delta)$ is a minimal *Dab*-formula at stage s iff, at stage s , $Dab(\Delta)$ is derived on condition \emptyset , and no $Dab(\Delta')$ with $\Delta' \subset \Delta$ is derived on condition \emptyset .

Definition 7 Where $Dab(\Delta_1), \dots, Dab(\Delta_n)$ are the minimal *Dab*-formulas derived on condition \emptyset at stage s , $U_s(\Gamma) = \Delta_1 \cup \dots \cup \Delta_n$.

Definition 8 Where Δ is the condition of line i , line i is marked at stage s iff $\Delta \cap U_s(\Gamma) \neq \emptyset$.

A formula A is said to be *derived at stage s* of a proof if it occurs on a line in the proof that is unmarked at stage s . As the marking proceeds in terms of the minimal *Dab*-formulas that are derived at a certain stage, it is clear that marking is a dynamic matter: a line may be unmarked at a stage s , marked at a later stage s' and again unmarked at an even later stage s'' . This is why a more stable notion of derivability is needed:

Definition 9 *A is finally derived from Γ at line i of a proof at stage s iff A is derived at line i at stage s , line i is unmarked at stage s , and every extension of the proof in which line i is marked has a further extension in which i is unmarked.*

As may be expected, the derivability relation of \mathbf{LA}_s^r is defined with respect to the notion of final derivability

Definition 10 $\Gamma \vdash_{\mathbf{LA}_s^r} A$ (*A is finally derivable from Γ) iff A is finally derived in an \mathbf{LA}_s^r -proof from Γ .*

The fact that \mathbf{LA}_s^r is in standard format immediately warrants that it has all interesting meta-theoretic properties, such as soundness and completeness and proof invariance.¹²

Theorem 1 $\Gamma \vdash_{\mathbf{LA}_s^r} A$ iff $\Gamma \vDash_{\mathbf{LA}_s^r} A$.

Theorem 2 *If $\Gamma \vdash_{\mathbf{LA}_s^r} A$, then every \mathbf{LA}_s^r -proof from Γ can be extended in such a way that A is finally derived in it.*

6 Conclusion and open problems

In this paper, I presented the logic \mathbf{LA}_s^r for the abduction of singular hypotheses. Despite some unusual properties (such as its dynamical character), \mathbf{LA}_s^r is a decent formal logic, with a nice proof theory and a (preferential) semantics that is sound and complete with respect to it. Several extensions and variants are possible for the logic \mathbf{LA}_s^r . One that first comes to mind is a variant that allows for the abduction of different explanatory hypotheses in cases where more than one *explanans* is available for the same *explanandum*. (In such cases, \mathbf{LA}_s^r only allows for the derivation of their disjunctions.) Other alternatives may be obtained by changing the extensional framework of \mathbf{LA}_s^r into one that allows, for instance, for a causal implication. It would also be interesting to explore alternatives in which different explanatory hypotheses do not necessarily have the same weight. Finally, it might be interesting to work on the heuristics behind abductive processes. The logic \mathbf{LA}_s^r simply determines which *explanantia* follow in view of an *explanandum* and a background theory. But, evidently, looking for an (interesting) explanation is a goal-directed process. All this goes beyond the scope of the logic presented in this paper. Still, it seems like an excellent starting point to explore these questions further.

¹²I refer to Diderik Batens, *Ibid.*, for an overview of the meta-theoretic properties and for their proofs.