

# An Adaptive Logic-based Approach to Abduction in AI\*

(Preliminary Report)

Tjerk Gauderis

Centre for Logic and Philosophy of Science  
Ghent University, Belgium

Tjerk.Gauderis@UGent.be

## Abstract

In a logic-based approach to abductive reasoning, the background knowledge is represented by a logical theory. A sentence  $\phi$  is then considered as an explanation for  $\omega$  if it satisfies some formal conditions. In general, the following three conditions are considered crucial: (1)  $\phi$  together with the background knowledge implies  $\omega$ ; (2)  $\phi$  is logically consistent with what is known; and (3)  $\phi$  is the most ‘parsimonious’ explanation. But, since abductive reasoning is a non-monotonic form of reasoning, each time the background knowledge is extended, the status of previously abduced explanations becomes once again undefined.

The adaptive logics program is developed to address these types of non-monotonic reasoning. In addition to deductive reasoning steps, it allows for direct implementation of defeasible reasoning steps, but it adds to each formula the explicit set of conditions that would defeat this formula. So, in an adaptive logic for abduction, a formula is an abduced hypothesis as long as none of its conditions is deduced. This implies that we will not have to recheck all hypotheses each time an extension to our background knowledge is made. This is the key advantage of this approach, which allows us to save repetitive re-computations in fast growing knowledge bases.

## 1 The Adaptive Logics Framework

The adaptive logics program is established to offer insight in the direct application of defeasible reasoning steps.<sup>1</sup> This is done by focussing on which formulas would falsify a defeasible reasoning step. Therefore, in adaptive logics a *formula* is a pair  $(A, \Delta)$  with  $A$  a regular well-formed formula in the

\*Research for this paper was supported by project subventions from the Special Fund for Research (BOF) of Ghent University. I am grateful to the anonymous referees for their helpful suggestions.

<sup>1</sup>The adaptive logics program is founded by Batens in the eighties. For a more recent overview of the general results, see [Batens, 2007]. For a philosophical defense of the use of adaptive logics, see [Batens, 2004].

language of the logic over which the considered theory  $\mathcal{T}$  is defined and  $\Delta$ , the *condition* of the formula, is a set of regular well-formed formulas that are assumed to be false. To express this assumption, these formulas are generally called *abnormalities* in adaptive logic literature.<sup>2</sup> For an adaptive logic in standard format, the abnormalities are characterized by a logical form.

The set of *plausibly derivable formulas*  $\mathcal{P}$  from a logical theory  $\mathcal{T}$  is formed in the following way:

1. *Premise Rule*: if  $A \in \mathcal{T}$ , then  $(A, \emptyset) \in \mathcal{P}$
2. *Unconditional Inference Rule*:  
if  $A_1, \dots, A_n \vdash B$   
and  $(A_1, \Delta_1), \dots, (A_n, \Delta_n) \in \mathcal{P}$ ,  
then  $(B, \Delta_1 \cup \dots \cup \Delta_n) \in \mathcal{P}$
3. *Conditional Inference Rule*:  
if  $A_1, \dots, A_n \vdash B \vee Dab(\Theta)$   
and  $(A_1, \Delta_1), \dots, (A_n, \Delta_n) \in \mathcal{P}$ ,  
then  $(B, \Delta_1 \cup \dots \cup \Delta_n \cup \Theta) \in \mathcal{P}$

where  $Dab(\Theta)$  stands for *disjunction of abnormalities*, i.e. the classical disjunction of all elements in the finite set of abnormalities  $\Theta$ . This third rule, which adds new conditions, makes clear how defeasible steps are modeled. The idea is that if we can deductively derive the disjunction of a defeasible result  $B$  and the formulas, the truth of which would make us to withdraw  $B$ , we can defeasibly derive  $B$  on the assumption that none of these formulas is true.

Apart from the set of plausible formulas  $\mathcal{P}$  we need a mechanism that selects which defeasible results should be withdrawn. This is done by defining a marking strategy. In the adaptive logics literature, several strategies have been developed, but for our purposes it is sufficient to consider the *simple strategy*. According to this strategy, the set of the *derivable formulas* or *consequences*  $\mathcal{D} \subseteq \mathcal{P}$  consists of :

1. *Deductive Results*: if  $(A, \emptyset) \in \mathcal{P}$ , then  $(A, \emptyset) \in \mathcal{D}$
2. *Unfalsified Defeasible Results*:  
if  $(A, \Theta) \in \mathcal{P}$  (with  $\Theta \neq \emptyset$ )  
and if for every  $\omega \in \Theta : (\omega, \emptyset) \notin \mathcal{P}$ ,  
then  $(A, \Theta) \in \mathcal{D}$

<sup>2</sup>This representation of adaptive logics is a reinterpretation of the standard representation of adaptive logics, which is in terms of a proof theory. I made this reinterpretation for purposes of comparison with other approaches in AI.

So, apart from the deductive results – which are always derivable – this strategy considers all defeasible results as derived, as long as none of the elements of their condition is deductively derived.

From the definitions of the sets  $\mathcal{P}$  and  $\mathcal{D}$ , we can understand how adaptive logics model the non-monotonic character of defeasible reasoning. If our theory  $\mathcal{T}$  is extended to the new theory  $\mathcal{T}'$  ( $\mathcal{T} \subset \mathcal{T}'$ ), then we can define the corresponding sets  $\mathcal{P}'$  and  $\mathcal{D}'$ . On the one hand, the set of plausibly derivable formulas will be monotonic ( $\mathcal{P} \subset \mathcal{P}'$ ), since there is no mechanism to withdraw elements from this set and it can only grow larger.<sup>3</sup> On the other hand, we know that the set of derivable formulas is non-monotonic ( $\mathcal{D} \not\subset \mathcal{D}'$ ). It is possible that a condition of a defeasible result in  $\mathcal{D}$ , is suddenly – in light of the new information in  $\mathcal{T}'$  – deductively derivable. So, this result will not be part of  $\mathcal{D}'$  any more. Obviously, no deductive result will ever be revoked.

This makes this kind of logics very apt to model fast growing knowledge bases.<sup>4</sup> If one needs a previously defeasibly derived result at a certain point, we cannot be sure whether it is still valid, because there might have been several knowledge base updates in the meantime. But, since the set of plausible formulas is monotonic, we know this formula will still be in  $\mathcal{P}$ . So, instead of recalculating the whole non-monotonic set  $\mathcal{D}$  after each knowledge base extension (which is the traditional approach), it is sufficient to expand the monotonic set  $\mathcal{P}$ . Of course, in this approach, if we want to use a defeasible result at a certain stage of knowledge base expansion, we will first have to check its condition. Still, it is easily seen that a lot of repetitive re-computation is avoided, certainly in situations in which we only need a small percentage of the defeasible results at every stage of knowledge base expansion.

Moreover, it is proven that if the adaptive logic is in standard format, which means that the abnormalities have a fixed logical form, the corresponding logic will have all interesting meta-theoretic properties. The logic for abduction developed in this article will be in standard format and will therefore be sound, complete, proof invariant and have the fixed-point property.<sup>5</sup>

## 2 Other “conditional” approaches

As far as I can see, two other approaches in AI have used the idea of directly adding conditions or restrictions to formulas. On the one hand, there is a line of research, called “Cumulative Default Reasoning”, going back to a paper of [Brewka, 1991] with the same title. On the other hand, in the area of argumentation theory, some work on defeasible logic programs (see, for instance, [García and Simari, 2004]) is also based on

<sup>3</sup>It is important to understand “plausible” as “initially plausible” (at the time of derivation) and not as “plausible according to our present insights”. The second definition would, of course, have led to a non-monotonic set.

<sup>4</sup>In that way, this kind of logic can offer a solution to what [Paul, 2000] mentioned as one of the main problems of both set-cover-based and some logic-based approaches to abduction.

<sup>5</sup>For an overview of the generic proofs of these properties, see [Batens, 2007].

formulas together with consistency conditions that need to be satisfied to make these formulas acceptable.

The main difference with these research programs is that the abnormalities in adaptive logics are based on a fixed logical form. This means that, for instance, the logical form for abduction – explained in this paper – is the form of abnormalities for any premise set on which we want to apply abductive reasoning. Put in other words, as soon as a fully classical premise set is given, all the possible abnormalities and, therefore, all the plausible and finally derivable abductive results can be calculated. There is no element of choice. In the other approaches, the conditions of defeasible steps must be given in the premise set, which leaves an element of choice which conditions we want to add to which defeasible implications. In adaptive logics, the defeasible consequences can be derived as soon as we have a classical premise set and as soon as we have chosen the appropriate logic for the kind of reasoning we want to do (e.g. abduction).

## 3 The problem of multiple explanatory hypotheses in Abduction

If we focus our attention now to the abductive problem, we cannot allow that the different defeasible results – the abducted hypotheses – are together in the set  $\mathcal{P}$ . For instance, if Tweety is a non-flying bird, he may be a penguin or an ostrich. But a set containing both the formulas ( $penguin(Tweety), \Theta_1$ ) and ( $ostrich(Tweety), \Theta_2$ ) is inconsistent.<sup>6</sup>

An elegant solution to this problem is found by translating this problem to a modal framework. When we introduce a possibility operator  $\diamond$  to indicate hypotheses and the corresponding necessity operator ( $\Box =_{df} \neg\diamond\neg$ ) to represent background knowledge, we evade this problem. The Tweety-example translates, for instance, as such (for variables ranging over the domain of all birds):

Background Knowledge:

$$\begin{aligned} &(\Box\forall x(penguin(x) \supset \neg flies(x)), \emptyset) \\ &(\Box\forall x(ostrich(x) \supset \neg flies(x)), \emptyset) \\ &(\Box\neg flies(Tweety), \emptyset) \end{aligned}$$

Plausible defeasible results:

$$\begin{aligned} &(\diamond penguin(Tweety), \Theta_1) \\ &(\diamond ostrich(Tweety), \Theta_2) \end{aligned}$$

So, with this addition the sets  $\mathcal{P}$  and  $\mathcal{D}$  are consistent again. Though, in this situation it is not really necessary to maintain the modal operators, because we can quite easily make a translation to a hierarchical set-approach, by borrowing some ideas of the Kripke-semantics for modal logics.<sup>7</sup> In these semantics, a hypothesis is said to be true in a possible world that is accessible from the world in which the hypothesis is stated, while necessities are true in all accessible worlds.

<sup>6</sup>At this point, we make abstraction of the exact conditions. The details of the conditions will be explained below.

<sup>7</sup>It is important to remember that we are constructing a syntactical representation, not a semantics for the underlying logic.

If we define now a world(-set) as the set of formulas assigned to that world, we can finish our translation from modalities to sets. We define the actual world  $w$  as the set of all formulas of the knowledge base and all deductive consequences. The elements of the set  $w$  are the only formulas that have a  $\Box$ -operator in our modal logic, and are thus the only elements that will be contained in every world-set in our system. Subsequently, for every abduced hypothesis we define a new world-set that contains it. This world is hierarchically directly beneath the world from which the formula is abduced. This new set contains further the formulas of all the world-sets hierarchically above, and will be closed under deduction. To make this hierarchy clear, we will use the names  $w_1, w_2, \dots$  for the worlds containing hypotheses directly abduced from the knowledge base,  $w_{1.1}, w_{1.2}, \dots, w_{2.1}, \dots$  for hypotheses abduced from a first-level world, etc.

With this translation in mind, we can omit the modal operators and just keep for every formula track of the hierarchically highest world-set that contains it. So, our Tweety example can be represented as such:

$$\begin{aligned} (\forall x(\text{penguin}(x) \supset \neg \text{flies}(x)), \emptyset) & w \\ (\forall x(\text{ostrich}(x) \supset \neg \text{flies}(x)), \emptyset) & w \\ (\neg \text{flies}(\text{Tweety}), \emptyset) & w \\ (\text{penguin}(\text{Tweety}), \Theta_1) & w_1 \\ (\text{ostrich}(\text{Tweety}), \Theta_2) & w_2 \end{aligned}$$

Since the hierarchical system of sets  $w_i$  is equivalent to the set  $\mathcal{P}$  (the plausibly derivable results) of a logic for abduction, the definition of the set  $\mathcal{D}$  (of this logic) can be applied to this system of sets too. It is clear that only the deductive consequences – the only formulas with an empty condition – will be the formulas in the set  $w$ . Further, since all formulas in a world-set have the same conditions, i.e. the condition of the hypothesis for which the world is created, the definition of  $\mathcal{D}$  does not only select on the level of the formulas, but actually also on the level of the world-sets.<sup>8</sup> Put in other words,  $\mathcal{D}$  selects a subsystem of the initial system of hierarchically ordered sets. The different sets in this subsystem are equivalent with what [Flach and Kakas, 2000] called *abductive extensions* of some theory. In this way, the logic can handle mutually contradictory hypotheses,<sup>9</sup> without the risk that any set of formulas turns out to be inconsistent.

#### 4 Reformulation of the abductive problem in the adaptive logics format

So far, in this paper we have shown – in the first section – how we can represent the standard format of adaptive logics in terms of two sets  $\mathcal{P}$  and  $\mathcal{D}$ , and – in the third section – how we can cope with contradictory hypotheses by using a hierarchical system of world-sets. In this section we will

<sup>8</sup>Strictly speaking, each world-set contains also all formulas of the world-sets hierarchically above. But since these formulas are also contained in those worlds above, no information is lost if we allow that  $\mathcal{D}$  can select on the level of the world-sets.

<sup>9</sup>Consider, for instance, the famous quaker/republican example: our approach will lead to two different abductive extensions, one in which Nixon will be a pacifist and another one in which he isn't.

now use this set representation to reformulate the syntax of the logic  $\text{MLA}^s$ , which is previously developed in [Gauderis, 2011].<sup>10</sup> This adaptive logic, the name of which stands for Modal Logic for Abduction, is an adaptive logic designed to handle contradictory hypotheses in abduction. The reformulation in terms of sets is performed with the goal to integrate the adaptive approach with other AI-approaches. First we need to define the abductive problem in a formal way.

**Definition 1.** An abductive system  $\mathcal{T}$  is a triple  $(\mathcal{H}, \mathcal{O}, d)$  of the following three sets

- a set of clauses  $\mathcal{H}$  of the form

$$\forall x(A_1(\alpha) \wedge \dots \wedge A_n(\alpha) \supset B(\alpha))$$

with  $A_1(\alpha), \dots, A_n(\alpha), B(\alpha)$  literals and  $\alpha$  ranging over  $d$ .

- a set of observations  $\mathcal{O}$  of the form  $C(\gamma)$  with  $C$  a literal and a constant  $\gamma \in d$ .
- a domain  $d$  of constants.

All formulas are closed formulas defined over a standard predicative first order logic.

Furthermore, the notation does not imply that predicates should be of rank 1. Predicates can have any rank, the only preliminaries are that in the clauses all  $A_i$  and  $B$  share a common variable, and that the observations have at least one variable that is replaced by a constant. Obviously, for predicates of higher rank, extra quantifiers for the other variables need to be added to make sure that all formulas are closed.

**Definition 2.** The background knowledge or actual world  $w$  of an abductive system  $\mathcal{T} = (\mathcal{H}, \mathcal{O}, d)$  is the set

$$w = \{(P, \emptyset) \mid \mathcal{H} \cup \mathcal{O} \vdash P\}$$

Since it was the goal of an adaptive logic-approach to implement directly defeasible reasoning steps, we will consider instances of the Peircean schema for abduction [Peirce, 1960, 5.171]:

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

When we translate his schema to the elements of  $\mathcal{T} = (\mathcal{H}, \mathcal{O}, d)$ , we get the following schema:

$$\frac{\forall x(A_1(\alpha) \wedge \dots \wedge A_n(\alpha) \supset B(\alpha)) \quad B(\gamma)}{A_1(\gamma) \wedge \dots \wedge A_n(\gamma)}$$

To implement this schema – better-known as the logical fallacy *Affirming the Consequent* – in an adaptive logic, we need to specify the logical form of the conditions that would falsify the application of this rule. As we can see from how the conditional inference rule is introduced in the first section, the disjunction of the hypothesis and all defeating conditions needs to be derivable from the theory. To specify these conditions, we will first overview the different *desiderata* for our abductions.

<sup>10</sup>In the original article, the syntax of the logic  $\text{MLA}^s$  is defined in terms of a proof theory.

Obviously, it is straightforward that if the negation of the hypothesis can be derived from our background knowledge, the abduction is falsified. If we know that Tweety lives in Africa, we know that he cannot be a penguin. So, in light of this information, the hypothesis cannot longer be considered as derivable:  $(penguin(Tweety), \Theta_1) \notin \mathcal{D}$ . But the hypothesis still remains in the monotonic set of ‘initially’ plausible results:  $(penguin(Tweety), \Theta_1) \in \mathcal{P}$ .

So, if we define  $A(\alpha)$  to denote the full conjunction,

$$A(\alpha) =_{def} A_1(\alpha) \wedge \dots \wedge A_n(\alpha)$$

the first formal condition that could falsify the defeasible step will be

$$\forall \alpha (A_1(\alpha) \wedge \dots \wedge A_n(\alpha) \supset B(\alpha)) \wedge B(\gamma) \wedge \neg A(\gamma).$$

To avoid self-explanations we will further add the condition that  $A(\alpha)$  and  $B(\alpha)$  share no predicates.

The reason why this condition also states the two premises of the abductive schema is because, in an adaptive logic, we can apply the conditional rule each time the disjunction is derivable. So, if we didn’t state the two premises in the abnormality, we could derive anything as a hypothesis since  $\vdash A(\gamma) \vee \neg A(\gamma)$  for any  $A(\gamma)$ . But with the current form, only hypotheses for which the two premises are true can be derived. This abnormality would already be sufficient to create an adaptive logic.

Still, we want to add some other defeating conditions. This could be done by replacing the abnormality by a disjunction of the already found abnormality and the other wanted conditions. Then, each time one of the conditions is derivable, the whole disjunction is derivable (by addition), and so, the formula defeated. But this result is obtained in the same way if we allow that one defeasible inference step adds more than one element to the condition instead of this complex disjunction. Hence, we will add these extra conditions in this way.

A lot of times, it is stated that the abduced hypothesis must be as parsimonious as possible. One of the main reasons for this is that one has to avoid random explanations. For instance, have a look at the following example:

$$\begin{aligned} \mathcal{H} &= \{\forall x (penguin(x) \supset \neg flies(x))\} \\ \mathcal{O} &= \{\neg flies(Tweety)\} \\ d &= \{x \mid x \text{ is a bird}\} \end{aligned}$$

The following formulas are derivable from this:

$$\begin{aligned} (\forall x (penguin(x) \wedge is\_green(x) \supset \neg flies(x)), \emptyset) & w \\ (penguin(Tweety) \wedge is\_green(x), \Theta_1) & w_1 \\ (is\_green(Tweety), \Theta_1) & w_1 \end{aligned}$$

The fact that *Tweety* is green is not an explanation for the fact that *Tweety* doesn’t fly, nor is it something that follows from our background knowledge. Since we want to avoid that our abductions yield this kind of random hypotheses, we will add a mechanism to control that our hypothesis is the most parsimonious.

A final condition that we have to add is that our observation is not a tautology. Since we use a material implication, anything could be derived as an explanation for a tautology, because  $\vdash B(\alpha) \supset \top$  for any  $B(\alpha)$ .

Now we can define the defeasible reasoning steps. Therefore we will need a new notation, which has the purpose to lift out one element from the conjunction  $A_1(\alpha) \wedge \dots \wedge A_n(\alpha)$ . This will be used to check for more parsimonious explanations.

**Notation 1** ( $A_i^{-1}(\alpha)$ ).

$$\begin{aligned} \text{if } n > 1 & : A_i^{-1}(\alpha) =_{df} (A_1(\alpha) \wedge \dots \wedge A_{i-1}(\alpha) \wedge \\ & A_{i+1}(\alpha) \wedge \dots \wedge A_n(\alpha)) \\ \text{if } n = 1 & : A_1^{-1}(\alpha) =_{df} \top \end{aligned}$$

**Definition 3.** The set of abnormalities  $\Omega$  for an abductive system  $T$  is given by

$$\begin{aligned} \Omega &= \{(\forall x (A_1(\alpha) \wedge \dots \wedge A_n(\alpha) \supset B(\alpha)) \wedge B(\gamma) \wedge \neg A(\gamma)) \\ & \vee \forall \alpha B(\alpha) \vee \bigvee_{i=1}^n \forall \alpha (A_i^{-1}(\alpha) \supset B(\alpha)) \mid \gamma \in d, \\ & \alpha \text{ ranging over } d, A_i \text{ and } B \text{ literals, } B \notin \{A_i\}\} \end{aligned}$$

It is easily seen that the generic conditional rule for adaptive logics – as defined in section 1 – defined by this set of abnormalities is equivalent with the following inference rule that is written in the style of the Peircean schema stated above.

**Definition 4.** Defeasible Inference rule for Abduction

$$\frac{\begin{array}{l} (\forall \alpha (A_1(\alpha) \wedge \dots \wedge A_n(\alpha) \supset B(\alpha)), \emptyset) \quad w \\ (B(\gamma), \emptyset) \quad w_i \\ (A_1(\gamma) \wedge \dots \wedge A_n(\gamma), \Theta) \quad w_{ij} \end{array}}{\quad}$$

with  $w_{ij}$  a new world hierarchically directly beneath  $w_i$  and  $\Theta = \{\neg(A_1(\gamma) \wedge \dots \wedge A_n(\gamma)), \forall \alpha B(\alpha), \forall \alpha (A_1^{-1}(\alpha) \supset B(\alpha)), \dots, \forall \alpha (A_n^{-1}(\alpha) \supset B(\alpha))\}$

So, it is possible to abduce further on hypothetical observations (and generate in that way further abductive extensions), but the implications need to be present in the background knowledge  $w$ . It is quite obvious, that if the abduced hypothesis is already abduced before (from, for instance, another implication), the resulting world-set will contain the same formulas, but with other conditions.

Finally, as explained in section 1, this body of definitions is formulated in the general framework of adaptive logics. This means that we have the following property.

**Property 1.** The logic  $MLA^s$  is a fixed-point logic which has a sound and complete semantics with respect to its syntax.

For the semantics and proof theory of this logic, and the proof that this logic is in the standard format of adaptive logics, we refer to [Gauderis, 2011]. For the soundness and completeness proof, we refer to the generic proof provided in [Batens, 2007] for all adaptive logics in standard format.

## 5 Example

**Motivation and comparison with other approaches** In this section we will consider an elaborate example of the dynamics of this framework. The main goal is to illustrate the key advantage of this approach, i.e. that there is no longer the need to recalculate all non-monotonic results at any stage of a growing knowledge base, but that one only needs to check

the non-monotonic derivability of the needed formulas at a certain stage against the monotonic plausibility.

This is the main difference with other approaches to abduction such as the ones explicated in, for instance, [Paul, 2000], [Flach and Kakas, 2000] or [Kakas and Denecker, 2002]. Since these approaches focus on a fixed and not an expanding knowledge base, they require in cases of expansion a full re-computation to keep the set of derived non-monotonic results updated. It is not claimed that the adaptive approach yields better results than these other approaches in cases of a fixed knowledge base. In fact, it is an issue for future research to investigate whether the integration of the existing approaches for fixed knowledge bases with the adaptive approach does not yield better results.

**Initial system  $\mathcal{T}$**  Our elaborate example will be an abductive learning situation about the observation of a non-flying bird, called Tweety. Initially, our abductive system  $\mathcal{T} = (\mathcal{H}, \mathcal{O}, d)$  contains in addition to this observation only very limited background knowledge.

$$\begin{aligned}\mathcal{H} &= \{\forall x(\text{penguin}(x) \supset \neg \text{flies}(x)), \\ &\quad \forall x(\text{ostrich}(x) \supset \neg \text{flies}(x))\} \\ \mathcal{O} &= \{\neg \text{flies}(\text{Tweety})\} \\ d &= \{x \mid x \text{ is a bird}\}\end{aligned}$$

Thus, our background knowledge contains the following formulas:

$$\begin{aligned}(\forall x(\text{penguin}(x) \supset \neg \text{flies}(x)), \emptyset) & \quad w & (1) \\ (\forall x(\text{ostrich}(x) \supset \neg \text{flies}(x)), \emptyset) & \quad w & (2) \\ (\neg \text{flies}(\text{Tweety}), \emptyset) & \quad w & (3)\end{aligned}$$

And the following abductive hypotheses can be derived:

$$\begin{aligned}(\text{penguin}(\text{Tweety}), \Theta_1) & \quad w_1 & (4) \\ (\text{ostrich}(\text{Tweety}), \Theta_2) & \quad w_2 & (5)\end{aligned}$$

with the sets  $\Theta_1$  and  $\Theta_2$  defined as

$$\begin{aligned}\Theta_1 &= \{\neg \text{penguin}(\text{Tweety}), \forall x \neg \text{flies}(x)\} \\ \Theta_2 &= \{\neg \text{ostrich}(\text{Tweety}), \forall x \neg \text{flies}(x)\}\end{aligned}$$

Since both implications have only one conjunct in the antecedent, their parsimony conditions – as defined in the general logical form – trivially coincide with the second condition. Since none of the conditions is deductively derivable in  $w$ , both (4) and (5) are elements of the set of derivable formulas  $\mathcal{D}$ .

**First Extension  $\mathcal{T}'$**  At this stage, we discover that Tweety can swim, something we know ostriches can't.

$$\begin{aligned}\mathcal{H}' &= \mathcal{H} \cup \{\forall x(\text{ostrich}(x) \supset \neg \text{swims}(x))\}, \\ \mathcal{O}' &= \mathcal{O} \cup \{\text{swims}(\text{Tweety})\} \\ d &= \{x \mid x \text{ is a bird}\}\end{aligned}$$

From which the following formulas can be derived:

$$\begin{aligned}(\forall x(\text{swims}(x) \supset \neg \text{ostrich}(x)), \emptyset) & \quad w & (6) \\ (\neg \text{ostrich}(\text{Tweety}), \emptyset) & \quad w & (7)\end{aligned}$$

Since the background information is extended, we only know that all previously derived hypotheses are still in the set of plausible hypotheses  $\mathcal{P}$ . If we want to check whether they are in the set of derivable hypotheses  $\mathcal{D}$ , we need to check whether their conditions are derivable from this extended information or not. But – this has already been cited several times as the key advantage of this system – we don't need to check all hypotheses. Since we don't have any further information on the penguin case, we just leave the hypothesis (4) for what it is. Thus, we save a computation, because at this stage we are not planning on reasoning or communicating on the penguin hypothesis. We just want to check whether this new information is a problem for the ostrich hypothesis; and indeed, it is easily seen that (5)  $\notin \mathcal{D}'$ .

**Second Extension  $\mathcal{T}''$**  At this stage, we will investigate further the penguin hypothesis and retrieve additional background information about penguins.

$$\begin{aligned}\mathcal{H}'' &= \mathcal{H}' \cup \{\forall x(\text{penguin}(x) \supset \text{eats\_fish}(x)), \\ &\quad \forall x(\text{on\_south\_pole}(x) \wedge \text{in\_wild}(x) \supset \text{penguin}(x))\} \\ \mathcal{O}'' &= \mathcal{O}' \\ d &= \{x \mid x \text{ is a bird}\}\end{aligned}$$

The following formulas can now further be retrieved:

$$(\text{eats\_fish}(\text{Tweety}), \Theta_1) \quad w_1 \quad (8)$$

$$(\text{on\_south\_pole}(\text{Tweety}), \Theta_{1.1}) \quad w_{1.1} \quad (9)$$

$$(\text{in\_wild}(\text{Tweety}), \Theta_{1.1}) \quad w_{1.1} \quad (10)$$

with the set  $\Theta_{1.1}$  defined as

$$\begin{aligned}\Theta_{1.1} &= \{\neg(\text{on\_south\_pole}(\text{Tweety}) \wedge \text{in\_wild}(\text{Tweety})), \\ &\quad \forall x \text{ penguin}(x), \\ &\quad \forall x(\text{on\_south\_pole}(x) \supset \text{penguin}(x)), \\ &\quad \forall x(\text{in\_wild}(x) \supset \text{penguin}(x))\}\end{aligned}$$

Since the first element of  $\Theta_{1.1}$  is actually a disjunction, the first condition can even be split in two.

This stage is added to illustrate the other aspects of adaptive reasoning. Firstly, as (8) illustrates, there is no problem in reasoning further on previously deductively derived hypotheses. Only, to reason further, we must first check the condition of these hypotheses (This poses no problem here, because we can easily verify that (4)  $\in \mathcal{D}''$ ). The deductively derived formula has the same conditions as the hypothesis on which it is built (and is contained in the same world). So, these results stand as long as the hypotheses on which assumption they are derived, hold. This characteristic of adaptive logics is very interesting, because it allows to derive predictions that can be tested in further investigation. In this example, we can test whether Tweety eats fish. In case this experiment fails and  $\neg \text{eats\_fish}(\text{Tweety})$  is added to the observations in the next stage, the hypothesis (and all results derived on its assumption) will be falsified. Secondly, the set of conditions  $\Theta_{1.1}$  for the formulas (9) and (10) contains now also conditions that check for parsimony. Let us illustrate their functioning with the final extension.

**Third Extension  $\mathcal{T}'''$**  At this stage, we learn that even in captivity the only birds that can survive on the South Pole are penguins. In addition to that, we get to know that Tweety is held in captivity.

$$\begin{aligned}\mathcal{H}''' &= \mathcal{H}'' \cup \{\forall x(\text{on\_south\_pole}(x) \supset \text{penguin}(x))\}, \\ \mathcal{O}''' &= \mathcal{O}'' \cup \{\neg \text{in\_wild}(\text{Tweety})\} \\ d &= \{x \mid x \text{ is a bird}\}\end{aligned}$$

If we now check the parsimony conditions of  $\Theta_{1.1}$ , we see that an element of this condition can be derived from our background knowledge. This means that all formulas assigned to world  $w_{1.1}$  are not derivable anymore on this condition. Still, one might wonder whether this parsimony condition should not keep (9) and only withdraw (10). But, that this is not a good road is proven by the fact that in that case (10) would be falsified by the extra observation that Tweety does not live in the wild. In fact, that it was a good decision to withdraw the whole world  $w_{1.1}$  is illustrated by the fact that the South Pole hypothesis of (9) can also be derived from  $\mathcal{H}'''$  in another world.

$$(\text{on\_south\_pole}(\text{Tweety}), \Theta_{1.2}) \quad w_{1.2} \quad (11)$$

with the set  $\Theta_{1.2}$  defined as

$$\Theta_{1.1} = \{\neg \text{on\_south\_pole}(\text{Tweety}), \forall x \text{penguin}(x)\}$$

So, at the end, we find that the set  $\mathcal{D}'''$  of derivable formulas consists of all formulas derivable in the worlds  $w$ ,  $w_1$  and  $w_{1.2}$ . The formulas of  $w_2$  and  $w_{1.1}$  are not an element of this final set of derivable results.

## 6 Conclusion

In this article we presented a new logic-based approach to abduction which is based on the adaptive logics program. The main advantages of this approach are :

1. Each abduced formula is presented together with the specific conditions that would defeat it. In that way, it is not necessary to check the whole system for consistency after each extension of the background knowledge. Only the formulas that are needed at a certain stage need to be checked. Furthermore, it allows for the conditions to contain additional requirements, such as parsimony.
2. In comparison with other approaches that add conditions to formulas, the conditions are here fixed by a logical form and hence only determined by the (classical) premise set. In this way, there is no element of choice in stating conditions (as, for instance, in default logics).
3. By integrating a hierarchical system of sets, it provides an intuitive representation of multiple hypotheses without causing conflicts between contradictory hypotheses.
4. It allows for further deductive and abductive reasoning on previous retrieved abduced hypotheses.
5. The approach is based on a proper sound and complete fixed point logic (MLA<sup>s</sup>).

**Limitations and Future Research** It has been argued that these advantages make this approach apt for systems in which not all non-monotonic derivable results are needed at every stage of expansion of a knowledge base. Still, it needs to be examined whether an integration with existing systems (for a fixed knowledge base) do not yield better results. Furthermore, since the key feature of this approach is the saving of computations in expanding knowledge bases, it needs to be investigated whether there is no integration possible with assumption-based Truth Maintenance Systems (building on the ideas of [Reiter and de Kleer, 1987]).

## References

- [Batens, 2004] Diderik 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.
- [Batens, 2007] Diderik Batens. A universal logic approach to adaptive logics. *Logica Universalis*, 1:221–242, 2007.
- [Brewka, 1991] Gerhard Brewka. Cumulative Default Logic. *Artificial Intelligence*, 50(2):183–205, 1991.
- [Flach and Kakas, 2000] Peter A. Flach and Antonis C. Kakas. Abductive and Inductive Reasoning: Background and Issues. In Peter A. Flach and Antonis C. Kakas, editors, *Abduction and Induction. Essays on their Relation and their Integration*, volume 18 of *Applied Logic Series*, pages 1–27. Kluwer Academic Publishers, Dordrecht, 2000.
- [García and Simari, 2004] Alejandro J. García and Guillermo R. Simari. Defeasible Logic Programming: An Argumentative Approach. *Theory and Practice of Logic Programming*, 4(1):95–2004, 2004.
- [Gauderis, 2011] Tjerk Gauderis. Modelling Abduction in Science by means of a Modal Adaptive Logic. *Foundations of Science*, 2011. Forthcoming.
- [Kakas and Denecker, 2002] Antonis Kakas and Marc Denecker. Abduction in Logic Programming. In A. Kakas and F. Sadri, editors, *Computational Logic: Logic Programming and Beyond. Part I*, pages 402–436. Springer Verlag, 2002.
- [Paul, 2000] Gabriele Paul. AI Approaches to Abduction. In Dov M. Gabbay and Rudolf Kruse, editors, *Abductive Reasoning and Uncertainty Management Systems*, volume 4 of *Handbook of Defeasible Reasoning and Uncertainty Management Systems*, pages 35–98. Kluwer Academic Publishers, Dordrecht, 2000.
- [Peirce, 1960] Charles S. Peirce. *Collected Papers*. Belknap Press of Harvard University Press, Cambridge, Massachusetts, 1960.
- [Reiter and de Kleer, 1987] Raymond Reiter and Johan de Kleer. Foundations of Assumption-based Truth Maintenance Systems: Preliminary Report. In *Proceedings of the Sixth National Conference on Artificial Intelligence (AAAI'87)*, pages 183–188, 1987.