# Using strict implication in background theories for abductive tasks

**M.G. Jansen** and **A.Th. Schreiber** and **B.J. Wielinga**
University of Amsterdam, Social Science Informatics
Roetersstraat 15, NL-1018 WB Amsterdam
E-mail jansen@swi.psy.uva.nl

## Abstract

Abduction is usually defined in terms of classical logical consequence. In this paper we substitute this 'inferential parameter' by the notion of *strict implication*. By doing so we hope to put more of the intended meaning of the abductive explanative relation into the background theory.

By using strict rather than material implication static domain knowledge for abductive tasks can deal with some limitations of the truthfunctional nature of material implication. It can be proved that by using strict instead of material implication the same explanations can be computed according to a strict version of abduction.

## 1 Introduction

In knowledge engineering it is common practice to separate the representation of static [1] domain knowledge from the procedural method for solving a knowledge intensive task. The reason for this is that such separation facilitates the reuse of the domain knowledge. When constructing domain models one has to realize that they might also be looked at outside the scope of the task they were designed for.

In abductive tasks the static domain knowledge is compiled in a logical theory, called the background theory. Abduction itself is often defined in terms of logical consequence and can therefore be considered a meta-logical notion. This meta-logical construct is not directly available at the object-level on which the background theory is formulated. As a result the background theory can be interpreted in two ways. First, since it is a classical logical theory it can be interpreted in the standard truth-functional way. Second, since it is input to an abductive procedure it has an 'explanative' interpretation. There is a discrepancy between the two interpretations.

When a domain model for an abductive task is written in propositional logic, the intended interpretation of this logical theory is the one in terms of possible explanations and observations. When the same theory is looked at outside the scope of the abductive task it has a standard truth-functional meaning.

[1] By 'static knowledge' is meant knowledge that will remain invariant during the problem solving process.

Object-level characterizations of abduction do exist [Console *et al.*, 1991] and show that the semantics of background theories is really that of completed theories, (in the sense of Clark-completion [Clark, 1978]).

Our approach is to leave the meta-logical nature of abductive inference intact, and instead to attempt to put more of the intended meaning on the object-level by changing the inferential parameter which occurs in the definition of abduction. By doing so we hope to solve some problems that may hamper the interpretation of domain theories outside the scope of the task. As an example of a different inferential parameter we will look at strict implication and define a notion of strict abduction accordingly.

This paper is structured as follows. First abduction is defined in general terms and the problems regarding the semantics of the domain knowledge are explained. The following section mentions a number of issues regarding the use of material implications that are at odds with the intended meaning of the representation. Next we introduce the notion of strict implication and define strict abduction in terms of it. It will be shown how strict implication deals with the aforementioned issues. A method for computing strict abduction is given in terms of analytical tableaux for T. Although this method is more complicated than that of a tableau method for classical abduction, for some theories the results are equivalent. Finally, a summary and general comments on our approach are given in the form of a discussion.

## 2 Abduction and representation

Abduction can be generally defined as follows [Aliseda-LLera, 1997]:

**Definition 1** *Given a logical theory* $\Theta$ *(a background theory) and a formula* $\beta$ *(an observation),* $\alpha$ *is an* abductive explanation *for* $\beta$ *iff the following properties hold:*

**Implication:** $\Theta, \alpha \Rightarrow \beta$

**Consistency:** $\Theta, \alpha \Rightarrow \beta$ *and* $\Theta, \alpha$ *is consistent.*

**Explanation:** $\Theta, \alpha \Rightarrow \beta$ *and* $\Theta \not\Rightarrow \beta$ *and* $\alpha \not\Rightarrow \beta$

**Minimality:** $\Theta, \alpha \Rightarrow \beta$ *and* $\alpha$ *is the weakest such explanation.*
*(*$\alpha$ *is the weakest explanation if* $\Theta, \alpha \Rightarrow \beta$ *and for all other formulas* $\phi$ *such that* $\Theta, \phi \Rightarrow \beta$*,* $\phi \rightarrow \alpha$*.)*

Here $\Rightarrow$ is a meta-logical symbol denoting an 'inferential parameter' [Aliseda, 2000] which relates the background theory and explanation to the observations. This can be classical or any other consequence relation, or any other 'implicative' relation. In the case that by the inferential parameter is meant *classical consequence* I will refer to the abductive variant as *classical abduction*.

It is important to note that this definition of abductive explanation is a meta-logical one. In fact the relation *abductive explanation* itself can be seen as a consequence relation, see for example Flach [Flach, 2000].

In an abductive task domain knowledge should somehow link explanations to observations. This static domain knowledge is represented in the background theory, which is part of the object-level language. The abductive explanation relation itself is not directly available at this level.

A consequence relation (for a logic $L$) is a meta-logical notion as well but it can be expressed on the object-level as implication, if the deduction theorem holds in $L$:

$$\Theta, \alpha \vdash_L \beta \Leftrightarrow \Theta \vdash_L \alpha \rightarrow \beta$$

As a result the domain knowledge necessary to link explanations to observations can be represented by the implication symbol of $L$. As a result, when compiling knowledge for background theories it is tempting to treat the implication symbol as the actual *abductive explanative* relation itself.

In the process of representing domain knowledge for an abductive task the use of implications is often taken to represent the meta-logical abductive explanative relation. But when we look at classical abduction the semantics of material implication does not reflect the intended meaning of its use. In fact there is a discrepancy between the meta-logical notion of abduction and the semantics of the object-level language.

## 3 Pragmatic issues

In order to illustrate that there is divergence between the object-level representation of the domain knowledge and the meta-logical construct of *abductive explanation* we will take a closer look at some issues regarding classical abduction and material implication.

These are inspired by pragmatic considerations about the representation of static domain knowledge. These issues also reflect the wish to put more of the intended meaning in the semantics of the object-level representation.

In the following subsections by 'abduction' 'classical abduction' is meant, by 'implication', 'material implication' and by 'consequence', 'classical consequence'.

### 3.1 Negation

By the above definition of abduction $\alpha$ is *not* an explanation if one of the mentioned properties fails to hold. For example, suppose for some $\alpha$ it is the case that $\alpha, \Theta \nvdash \beta$. To express this explicitly in the background theory cannot be done by simply negating the implication $\alpha \rightarrow \beta$. The reason for this that $\neg(\alpha \rightarrow \beta) \vdash \alpha \wedge \neg\beta$, which is clearly not what is intended.

However one could view the background theory as a closed world in the sense that $\alpha$ is *not* an explanation for $\beta$ if it cannot be proved that $\alpha$ is an explanation of $\beta$. In such theories every abducible is either an explanation of an observed phenomenon or not.

Formally we can define this as follows:

**Definition 2** Completeness assumption for background theories: *Given $\Theta$ (a background theory) and $\beta$ (an observation), $\alpha$ is* not *an abductive explanation for $\beta$ iff $\Theta, \alpha \nvDash \beta$ and $\Theta \cup \alpha$ is consistent.*

The idea that $\alpha$ is not an explanation of $\beta$ by failure to proof that it is, might seem overconstrained. For many explanative models this completion assumption seems to be too strong. Console et al. [Console and Torasso, 1990] circumvented this objection by the use of *incompleteness-assumption symbols*. Let $A$ be such a symbol, then $A \rightarrow \beta$ expresses that $A$ is an unknown explanation of $\beta$. This construction leaves room for explicitly stating that the logically completed theory is not complete in a pragmatic sense.

### 3.2 Problems with conditionals

There are a number of problems regarding the use of material implication in conditionals which are well documented (for an overview see [Veltman, 1985]). The main issue that concerns us here is that the truth-functional meaning of conditional statements may lead to confusion when they are used to represent background knowledge for abductive tasks.

Any (material) conditional statement $\alpha \rightarrow \beta$ can be reformulated as $\neg\alpha \vee \beta$. This last formulation makes an interpretation of the implication as *explains* cumbersome.

The conditional problems with material implication are perhaps best illustrated by what are known as the 'paradoxes' of material implication: Consider the following two sentences, both theorems of propositional calculus (PC):

**(1)** $p \rightarrow (q \rightarrow p)$

**(2)** $\neg p \rightarrow (p \rightarrow q)$

The first one can be said to mean that if a proposition is true any proposition implies it. The sense of the second is that anything is implied by a proposition which is false. From these it can be derived that:

**(3)** $(p \rightarrow q) \vee (q \rightarrow p)$

Calling this a paradox might suggest that this is a fallacy but it is not. The above sentences merely reflect the truth-functional meaning of material implication.

From these it follows that the semantics of material implication is indeed different from that of *abductive explanation*. For example theorem (3) would say that for every two phenomena one is the explanation of the other. The contrary, that two phenomena are each not to be considered as an explanation of the other, can therefore not be expressed. Clearly, this is not in line with the definition of abductive explanation.

### 3.3 Necessity and possibility

In more complex domains often expressions are needed which express the necessity or possibility of explanative relations. Classical logic offers no standard way of modelling constructs like *A possibly causes B* and *A necessarily causes*

*B*. Several solutions using classical logic have been proposed though.

Poole [Poole, 1988] advocates an approach where possible and necessary causal implications are represented by two different sets. One, named H, contains general or open hypotheses and the other F, closed formula or facts. The interpretation is that H contains implications which should be interpreted as possible, whereas elements of F should be interpreted as necessary implications.

Another approach is the above described use of assumption symbols in Console *et al*[Console and Torasso, 1990]. Here the fact that *A MAY cause B* is represented as $A \vee \alpha \rightarrow B$ where $\alpha$ stands for an hypothetical assumption the truth of which is first assumed but which may be rejected if their evaluation gives rise to unwanted results.

However, the modalities *possibility* and *necessity* are not directly available at the object-level. Using a modal instead of a truth-functional semantics would solve this issue.

### 3.4 Non-deterministic explanation

By an non-deterministic explanation is meant here that some phenomenon $\alpha$ can explain observations $\beta$ *or* $\gamma$. As $\alpha$ occurs $\beta$ or $\gamma$ will result but it isn't known which one. The non-deterministic aspect indicates a choice between $\beta$ and $\gamma$.

When representing knowledge for an abductive task, one could make use of a conditional of the form:

$$\alpha \rightarrow \beta \vee \gamma$$

However it is not clear what such a conditional should mean. If we take the viewpoint that implication is interpreted as an explanative relation its intended meaning would be something like: $\alpha$ *explains* $\beta$ *or* $\gamma$.

However when performing abduction and $\beta$ is observed this conditional will not lead to the conclusion that $\alpha$ is an explanation.

$$\alpha \rightarrow \beta \vee \gamma, \alpha \vdash \beta \vee \gamma$$

And $\beta \vee \gamma$ does not imply $\beta$.

This problem occurs frequently in classification tasks. Consider the following representation of the fact that a blackbird is either black or brown:

$blackbird \rightarrow (colour = brown \vee coulour = black)$

The attribute $colour$ has as possible values $\{black, brown\}$. Observations are represented as attribute-value pairs where only one value per attribute is allowed. Therefore if the observation $colour = black$ is made, it follows that $colour \neq brown$.

At first sight $blackbird$ seems to cover this observation, but closer inspection learns that it does not. Though the observation is consistent with $blackbird$, it is not implied by it.

## 4 An alternative approach

The issues raised in the previous section were meant to draw attention to the fact that the semantics of the object-level language in which the background theory is formulated does not reflect the intended meaning of the abductive explanative relation.

One solution is to approach the notion of abductive explanation as an object-level notion itself. This has been described by Console *et al.* [Console *et al.*, 1991] and Konolige [Konolige, 1992]. From the first [Console *et al.*, 1991] it has become clear that for cycle-free background theories which contain only definite clauses, their semantics in an abductive setting is really that of *completed* theories.

The problems raised in the previous section can be dealt with in this manner. For example, it becomes clear that to express explicitly in a background theory that some $\alpha$ is *not* to be regarded as an explanation, has a strong resemblance to negation as failure. Possibility can be expressed in completed theories as well, see Console *et al*[Console and Torasso, 1990].

Here we take a different approach. Instead of bringing the notion of abduction to the object-level we try to bridge the gap between observation and explanation by looking at another logical system. So instead of substituting classical consequence for the inferential parameter in the general definition of abduction given above, others can be tried.

There is however one potential drawback to this approach. A decision procedure for an inferential parameter $\Rightarrow$ can be adapted in order to compute abduction with. (An example method in the form of analytical tableaux will be described later.) Another notion of logical consequence in the definition of abduction, will generally lead to a different method for computing abductive explanations. In order to avoid complicating this procedure care must be taken when choosing an inferential parameter. If the resulting logic has no or an arduous decision procedure this will complicate the abduction procedure considerably.

Here we will describe one alternative inferential parameter: a system for *strict implication*, (strict implication will be denoted by the symbol $⥽$). It is important to note that if in a propositional logic we replace every occurrence of $\rightarrow$ by $⥽$ the result is a weaker logic, in the sense that every theorem of the latter is also one of the former (but not *vice versa*).

Strict implication is certainly not the only candidate for our purposes. Another attempt could consist of formulating abduction in terms of intuitionistic logic. We will leave this for future research.

## 5 Strict implication

The notion of strict implication was first put forward in modern times by Lewis, [Lewis and Langford, 1959]. The intention was to come up with a different notion of implication which would not lead to the paradoxes of material implication, mentioned above.

The intuitive meaning of strict implication can be formulated as: A sentence of the form $\alpha ⥽ \beta$ is true in a given situation $s$ iff there is no possible situation $s'$ such that $\alpha$ is true in $s'$ and $\beta$ is false in $s'$. Lewis' distinction between *strict* (or *necessary*) and *material* implication marked the birth of the development of modern modal logic.

In modern notation strict implication ($⥽$) can be defined as $\Box(\alpha \rightarrow \beta)$, or alternatively $\neg\Diamond(\alpha \wedge \neg\beta)$. [2]

The modal operators $\Box$ (necessity) and its dual $\Diamond$ (possibility) now facilitate to express directly the distinctions between three kinds of propositions: The tautologies or necessary true

[2] $\Box\alpha$ can be defined as $\neg\Diamond\neg\alpha$. Similarly $\Diamond$ can be defined in terms of $\Box$.

propositions (represented by $\Box\alpha$), the contradictions or impossible propositions (represented by either $\Box\neg\alpha$ or $\neg\Diamond\neg\alpha$) and the contingencies which are neither contradictions nor tautologies (represented by either propositional variables or $\Diamond\alpha$.

The weakest system to capture these ideas is the modal logic T. [3] Stronger systems like S4 and S5 could be used but these are extensions of T. Furthermore the decision procedure for a logic can often be adapted in order to perform abduction with. For example [Aliseda-LLera, 1997] uses semantic tableaux. Modal tableaux procedures for T and S4 exist [Fitting, 1983; Gore, 1999] but not for S5. Not surprisingly the procedure for S4 is more complicated than the one for T. This general tradeoff between complexity and expressiveness is the main motivation to opt for a weak alternative for classical abduction.

However, as said above just replacing material by strict implication yields a weaker propositional logic, whereas T is stronger than classical propositional logic: it contains PC. The reason for choosing T in preference to the weaker strict implication logics, is that the latter do not come with a decision procedure, and so computing abduction in them is not evident.

Syntactically T contains PC and in addition has the following two axioms:

**(K)** $\Box(\alpha \rightarrow \beta) \rightarrow (\Box\alpha \rightarrow \Box\beta)$

**(T)** $\Box\alpha \rightarrow \alpha$

and the following extra rule:

**(Necessitation)** If $\alpha$ is a theorem of T so is $\Box\alpha$.

The deduction theorem, mentioned above, holds for T.

The semantics of the modal logic T can be characterized by two key notions: *possible worlds* and *accessibility relation*. A Kripke-model is built up from several possible worlds, representing as many possible states of affairs. In these models truth-values are always connected to possible worlds rather than formulas. In determining the truth-value of a purely propositional formula in a world other worlds play no role. Only if $\Box$ occurs may it be necessary to involve other worlds.

Possible worlds are connected by means of an *accessibility relation* which together constitute a frame. Worlds may be accessible from or visible for each other. A formula $\Box\alpha$ is true in world $w$ if $\alpha$ is true in all worlds $w'$ accessible from $w$. A formula is valid if it is true under any interpretation in every world.

The semantics of T differs from that of other modal systems in that its frames (characterized by the T-axiom) are *reflexive*. For a more formal treatment of modal logic semantics the reader is referred to [Hughes and Creswell, 1996].

The notion of strict implication makes use of the modality *necessity*. Hence, using strict implication instead of material implication means making use of a modal instead of a truth-functional semantics.

[3] Lewis originally came up with weaker systems S1 and S2. These do not make use of the modalities of necessity and possibility. T is the weakest *normal* system. For a discussion of Lewis' systems and the appropriateness of T, the interested reader is referred to Hughes & Creswell [Hughes and Creswell, 1996].

## 6 Strict abduction

The purpose of this exercise is to see if we can use strict instead of material implication in abductive background theories. In order to do this we define the notion of *strict abduction*. This comes down to substituting strict implication for the inferential parameter in the general definition of abduction described in section 2. In addition it should be clear that all reasoning should be done in the system T.

**Definition 3** *Given $\Theta$ (a background theory) and $\beta$ (an observation), $\alpha$ is a* strict *abductive explanation for $\beta$ in the system T iff:*

**Implication:** $\Theta \vdash_T \Box(\alpha \rightarrow \beta)$

*The rest of the properties can be defined in a similar way.*

Note that strict abduction is *not* the same as abduction in $T$. $T$ contains PC and performing abduction in it would lead to a more complicated procedure for computing abductive explanations. Just replacing strict for material implication in a background theory does not lead to a more complicated procedure, as will be shown below.

Let $\Theta$ be a purely propositional (consistent) background theory consisting of purely propositional implications of the form $\alpha \rightarrow \beta$. $\alpha$ is here an abducible atom, a disjunction or conjunction of atoms.

Let $\Box\Theta$ be the set [4] where each material implication in $\Theta$ has been replaced by a strict implication. Furthermore let $\beta$ be an observation. Now the following holds:

**Proposition 1** $\alpha$ is a *classic* abductive explanation for $\beta$ according to $\Theta$ iff $\alpha$ is a *strict* abductive explanation for $\beta$ according to $\Box\Theta$.

**Proof:** It suffices to show that
$\vdash \Theta \rightarrow (\alpha \rightarrow \beta) \Leftrightarrow \vdash_T \Box\Theta \rightarrow \Box(\alpha \rightarrow \beta)$

**($\Rightarrow$)** Since T contains PC: $\vdash_T \Theta \rightarrow (\alpha \rightarrow \beta)$ then by necessitation and the K-axiom it follows that $\vdash_T \Box\Theta \rightarrow \Box(\alpha \rightarrow \beta)$.

**($\Leftarrow$)** This proof depends on the fact that each thesis of $T$ has a PC-transform which is valid in PC. The PC-transform of a formula $\alpha$ is formed by rewriting it in a form containing only $\neg$, $\wedge$, $\vee$ and $\Box$ and then removing every occurrence of $\Box$. For the proof see [Hughes and Creswell, 1968]. It can be verified that the PC-transform of $\Box\Theta \rightarrow \Box(\alpha \rightarrow \beta)$ is $\Theta \rightarrow (\alpha \rightarrow \beta)$.

This means that there is no difference between strict and classic abduction if the only difference between the domain theories is that the strict one contains boxed versions of the classic one.

In fact it also shows that in background theories for classical abductive tasks the implications which occur in the background theory *can* be safely interpreted as strict implications. They can also be interpreted as possible implications.

When it is clear that the implications in the background theory should be interpreted as necessary implications there is no need for a special procedure for strict abduction.

[4] Although the domain theories are defined as sets, they are sometimes used as a conjunction of their elements. Note that $\Box$ can be distributed over conjunctions: $\Box(\alpha \wedge \beta) \Rightarrow \Box\alpha \wedge \Box\beta$.

However, when combining the necessary and possible explanations such a procedure is needed. We will describe such a method in terms of analytical tableaux in T. As an introduction to these ideas we first give a quick introduction into the method of analytical tableaux for propositional logic, and propositional classical abduction. Those familiar with the tableau method can skip the next section.

## 7 Analytical tableaux for abduction

Decision procedures for classical logic like resolution and analytical tableaux can be used in a 'reverse manner' for abductive reasoning. To illustrate the ideas presented in this paper analytical tableaux will be used. The choice over other methods is purely a pragmatic one. Tableaux are a quite popular method in automated theorem proving [Fitting, 1990] and their application in the field of abduction is well-documented. [Aliseda-LLera, 1997] [Mayer and Pirri, 1993].

In a tableau proof a set of formulae is transformed into binary tree by means of reduction rules. These rules are depicted in table 1.

$$\frac{\neg\neg Z}{Z} \qquad \frac{\neg\top}{\bot} \qquad \frac{\neg\bot}{\top} \qquad \frac{\alpha}{\begin{matrix}\alpha^1\\ \alpha^2\end{matrix}} \qquad \frac{\beta}{\beta^1 \mid \beta^2}$$

Table 1: Rules for the tableau trees

The first rule indicates that double negations are redundant. The second and third rule deal with the atomic constants $\top$ and $\bot$. All propositional formulas containing binary connectives can be divided as belonging to two types: True conjuctive formulas ($\alpha$-type) and true disjunctive formulas ($\beta$-type) [Fitting, 1990]. The rule for $\alpha$-type formulas indicates that the conjuncts have to be placed on the same branch of the tree. The $\beta$-rule however indicates a branching of the tree.

If at any branch a formula and its negation appear, the branch is said to be closed. If all branches close, the tree is said to be closed. A tableau tree $T(\Theta)$ for a theory $\Theta$ has the following two general logical properties: Every open branch in a tableau $T(\Theta)$ corresponds to a verifying model. If $T(\Theta)$ is closed, $\Theta$ is inconsistent.

In order to test whether a certain formula $\alpha$ follows from a set of premisses $\Theta$ a tableau tree is constructed for $\Theta \cup \{\neg\alpha\}$. Only if the constructed tableau closes does $\alpha$ follow from $\Theta$.

The procedure for performing plain abduction with tableaux is as follows: The tableaux $T(\Theta \cup \{\neg\beta\})$ is generated. Assuming that $\beta$ does not follow from $\Theta$ alone, this results in an open tableau. Any formula $\alpha$ which closes the tableau when added to it is an explanation.

The various styles of abduction can then be described as follows: Let $\Theta$ be the domain theory, $\beta$ the observation, then $\alpha$ is an explanation if:

**Plain:** $T((\Theta \cup \{\neg\beta\}) \cup \{\alpha\}$ is closed

**Consistent:** Plain +
$T(\Theta \cup \{\alpha\})$ is open.

**Explanatory:** Plain +
$T(\Theta \cup \{\neg\beta\})$ is open and $T(\alpha \cup \{\neg\beta\})$ is open.

**Minimal:** Plain +
$\alpha$ is minimal.

In addition abducibles should be part of the vocabulary of the domain theory and the observation, and be either literals or conjunctions or disjunctions of literals. For minimal abduction literals should be checked first. If none of these close the tableau conjunctions of literals should be checked for, etc.

## 8 Strict implication in T

Tableaux in T (or T-tableaux) can be constructed with the help of the reduction rules for PC and two additional rules for the modal operators, shown in table 2. [5]

$$\frac{\Diamond\neg\alpha, \Box\Sigma, \Sigma_1}{\neg\alpha, \Sigma} \qquad \frac{\Box\alpha, \Sigma}{\Box\alpha, \alpha, \Sigma}$$

Table 2: Additional rules for T-tableaux.

Where $\Box\Sigma$ is a set of boxed formulae, $\Sigma_1$ a set of non-boxed formulae and $\alpha, \Sigma, \Sigma_1$ is short for $\{\alpha\} \cup \Sigma \cup \Sigma_1$.

The interpretation of these rules is different from those for PC. After application the formulae on the top side of the line must be deleted from the current branch. The reason for this is that the two modal rules mimic the transition to another world. Boxed formulae are then stripped from a box and unboxed formulae disappear, as they are local with respect to the world they occur in.

With the help of semantic tableaux for T strict abduction can now be performed. Let $\Theta$ be the domain-theory, $\beta$ the observation and $\alpha$ an explanation. The goal is to derive $\Box(\alpha \to \beta)$ in T for some formula $\alpha$. Using tableaux this can be done as follows:
Generate the T-tableau $T(\Theta \cup \{\neg\Box(\alpha \to \beta\})$
Check which formulae when substituted for $\alpha$ result in closure of T.

Apart from the fact that T-tableaux instead of PC-tableaux are used, a procedure for strict abduction differs in two ways from a normal abductive procedure. First, not the negation of the observation is added to the tableau of the domain theory, but instead the negated strict implication from explanation to observation. Second, explanations are not those formulae which close the tableau when added to it, but when substituted for $\alpha$.

## 9 Issues revisited

If instead of material implication strict implication is used the issues mentioned before can be treated directly in the language of the background theory.

**Negation.** To express that $\alpha$ does not explain $\beta$ can be expressed by $\neg\Box(\alpha \to \beta)$. In fact the semantics of this statement compared to the material conditional $\neg(\alpha \to \beta)$ is much closer to the intended meaning. As the last conditional means $\alpha$ and not $\beta$, the strict version has the meaning that it is *possible* that $\alpha$ and not $\beta$.

[5] A more elaborate description of tableaux for modal systems can be found in [Fitting, 1983] and [Gore, 1999].

Note that if $\neg\Box(\alpha \to \beta)$ is contained in the background theory $\alpha$ is indeed not an explanation for $\beta$ since the implication property in the definition of strict abduction does not hold.

**Conditional problems.** The notion of strict implication was developed out of dissatisfaction with the paradoxes of material implication. These do not hold for strict implication. In that sense strict implication is to be preferred over material implication. However, strict implication is not the ideal candidate for solving all problems concerning conditional statements.

Like material implication $\mathrel{-\!\!3}$ is still a connective and does not really express the *connection* between explanation and observation. This objection can be put in the form of 'paradoxes of strict implication':

$$(p \wedge \neg p) \mathrel{-\!\!3} q$$
$$q \mathrel{-\!\!3} (p \vee \neg p)$$
$$\Box\neg p \to (p \mathrel{-\!\!3} q)$$
$$\Box q \to (p \mathrel{-\!\!3} q)$$

All these sentences are theorems of T. The meaning of each of them can be expressed respectively as follows: a contradiction strictly implies everything. Every proposition strictly implies a tautology. Every impossible proposition strictly implies everything. When a proposition is necessary it is strictly implied by everything. Similar to the paradoxes of material implication these sentences may give rise to unwanted interpretations.

**Possibility and necessity** The modal semantics of T now facilitates expressing possibility and necessity directly in the object-level language. However, this would mean that the notion of abduction should be adapted in order to make the distinction between possible and necessary (strict) explanations.

In fact *possible* abduction can easily be defined as strict implication, except that each occurrence of $\Box$ is replaced by $\Diamond$. For possible abduction a similar result as that of proposition 1 can be proved.

**Proposition 2** $\alpha$ is a classic abductive explanation for $\beta$ according to $\Theta$ iff $\alpha$ is a possible abductive explanation for $\beta$ according to $\Diamond\Theta$.

**(⇒)** $\Theta \to (\alpha \to \beta)$ holds in PC and therefore in T. So:
$\vdash_T \neg(\alpha \to \beta) \to \neg\Theta \;\Leftrightarrow$
$\vdash_T \Box\neg\Theta \to \Box\neg(\alpha \to \beta) \;\Leftrightarrow$
$\vdash_T \neg\Box\neg(\alpha \to \beta) \to \neg\Box\neg\Theta \;\Leftrightarrow$
$\vdash_T \Diamond(\alpha \to \beta) \to \Diamond\Theta$ [6]

**(⇐)** Since $\Diamond$ can be defined as $\neg\Box\neg$, the PC-transform is the same as that of the boxed version.

This result shows that material implication can be interpreted in two ways: As possible or necessary explanation. Using both at the same time would mean performing strict abduction in T. This would lead to the more complicated abductive procedure (described above) than the standard case.

[6] Note that $\Diamond$ can be distributed over conjunctions: $\Diamond(\alpha \wedge \beta) \Rightarrow \Diamond\alpha \wedge \Diamond\beta$.

**Non-determinism and abduction** Suppose $\alpha$ explains either $\beta$ or $\gamma$. If $\beta$ is observed will an abductive procedure produce $\alpha$ as an explanation?

The answer partly depends on the representation of the explanative rule. Just using material implication would give us $\alpha \to (\beta \vee \gamma)$. Then clearly $\alpha$ does not imply $\beta$. However one could opt for a *weak* variant of abduction where $\alpha$ is an explanation only if it is consistent with the observation. Although this would give the desired result, this procedure will often lead to numerous unwanted explanations as well.

Another problem with such a weak variant of abduction is that the relevance of the explanation for the observation is still lower than for classic abduction.

Strict implication does not provide for an ideal solution to this problem. $\alpha$ is not a strict explanation for $\beta$ if the explanative rule is represented as $\Box(\alpha \to (\beta \vee \gamma))$.

However making a distinction between possible and necessary causal relations does offer a solution. Non-determinism expresses a choice, or possibility, between alternatives. The non-deterministic causal rule could thus better be represented as: $\Diamond(\alpha \to \beta) \wedge \Diamond(\alpha \to \gamma)$.

Using this representation $\alpha$ is not a *strict* but a possible explanation for $\beta$.

However this solution still does not completely explain some cases. Consider the example where one disease has as symptoms red skin and another either skin-rash or fever. Observing skin-rash would mean the first disease to be a *necessary*, the second a *possible* explanation. Since possibility is the weaker notion this suggests that the the necessary one is to be preferred. The question remains if this is really the desired result.

## 10 Discussion

By using strict implication and defining abduction in terms of it, the semantics of the object-level representation changes from a truth-functional to a modal one. Still the same explanations can be computed on a *strict* background theory compared to the classical case. The interpretation of background theories in terms of strict implication has a number of advantages in that it deals with some pragmatic issues regarding the intended meaning of the knowledge representation.

We do not claim that strict implication is the only notion that could be used to solve problems like the ones discussed. In fact intuitionistic logic seems to be a good candidate as well. The semantics of this logic can be formulated in terms of information states, which is interesting from an abductive point of view. Furthermore this logic is weaker than classical logic.

The discrepancy between the semantics of the representation of static domain knowledge and meta-level construct is of interest for the field of knowledge representation. Abduction can be seen as a (general) problem solving method operating upon static knowledge representations. As such it interprets the domain knowledge in a partial way. As a result this 'procedural semantics' of the domain knowledge differs from the semantics of the representation language itself.

## References


[Aliseda-LLera, 1997] A. Aliseda-LLera. *Seeking explanations: Abduction in Logic, Philosophy of science and Artificial Intelligence*. PhD thesis, ILLC, University of Amsterdam, Amsterdam, The Netherlands, 1997.

[Aliseda, 2000] A. Aliseda. *Abduction as epistemic change*, volume 18 of *Applied Logic Series*, chapter 3, pages 45–58. Kluwer Academic Publishers, Dordrecht, 2000.

[Clark, 1978] K. L. Clark. On closed-world data bases. *Logic and Databases*, 1978.

[Console and Torasso, 1990] L. Console and P. Torasso. Hypothetical reasoning in causal models. *Int. J. of Intelligent Systems*, 5(1):83–124, 1990.

[Console *et al.*, 1991] L. Console, D.T. Dupre, and P. Torasso. On the relationship between abduction and deduction. *Journal of Logic and Computation*, 1(5):661–690, 1991.

[Fitting, 1983] M. Fitting. *Proof methods for modal and intuitionistic logics*. D. Reidel Publishing Company, Dordrecht-Boston-Lancaster, 1983.

[Fitting, 1990] M. C. Fitting. *First-Order Logic and Automated Theorem Proving*. Springer-Verlag, New York, 1990.

[Flach, 2000] P.A. Flach. *On the logic of hypothesis generation*, volume 18 of *Applied Logic Series*, chapter 6, pages 89–106. Kluwer Academic Publishers, Dordrecht, 2000.

[Gore, 1999] R. Gore. Tableau methods for modal and temperal logics. In M. D'Agostino, D. Gabbay, R. Hahnle, and J. Posegga, editors, *Handbook of Tableau Methods*. Kluwer, Dordrecht, 1999.

[Hughes and Creswell, 1968] G. E. Hughes and M.J. Creswell. *An introduction to modal logic*. Methuen, London and New York, 1968.

[Hughes and Creswell, 1996] G. E. Hughes and M.J. Creswell. *A new introduduction to modal logic*. Routledge, London and New York, 1996.

[Konolige, 1992] Kurt Konolige. Abduction versus closure in causal theories. *AI Journal*, 53:255–272, 1992.

[Lewis and Langford, 1959] C.I. Lewis and C.H. Langford. *Symbolic logic*. Dover publications, New York, 1959.

[Mayer and Pirri, 1993] M. Mayer and F. Pirri. First order abduction via tableau and sequent calculi. *Bulletin of the IGPL*, 1:99–117, 1993.

[Poole, 1988] D. L. Poole. Representing knowledge for logic-based diagnosis. In *Proc. International Conference on Fifth Generation Computing Systems*, pages 1982–1290, Tokyo, 1988.

[Veltman, 1985] F Veltman. *Logic for conditionals*. PhD thesis, University of Amsterdam, Amsterdam, The Netherlands, 1985.