A Logical Framework for Abduction

Maurice Pagnucco
email: morri@cs.su.oz.au

Norman Y Foo
email: norman@cs.su.oz.au

Abhaya Nayak
email: abhaya@cs.su.oz.au

Knowledge Systems Group
Basser Department of Computer Science
Madsen Building F09
University of Sydney
NSW, 2006, Australia
fax : +61-2-692-3838
tel : +61-2-692-3423

Category: Paper.

Abstract

Abductive reasoning is a fundamental type of logical inference that attempts to determine explanations for data. The aim of this paper is to investigate the notion of abduction and, in particular, various types of abduction within a general logical setting.

We begin with a standard definition of abduction commonly found in artificial intelligence literature, modify this slightly and proceed to define particular types of abduction which characterise important aspects of abductive reasoning. The definitions are expressed in terms of a consequence relation $\vdash$ which is assumed to include classical propositional logic and obey compactness and the deduction theorem.

We finally discuss two "degenerate" types of abduction and their relationships with other reasoning processes.

Keywords: Abductive reasoning, logical inference, non-standard inference.

A Logical Framework for Abduction

1 Introduction

Abductive reasoning (or abduction) is a form of logical inference that aims to derive "explanations" for a body of data. In fact, it is often described as "inference to the best explanation" [7]. The term "abduction" was coined by the American philosopher Charles Sanders Peirce though its origins can be traced back to Aristotle. Peirce was the first person to distinguish it as a fundamental type of logical inference alongside deduction and induction [5].

The notion of abduction is, in essence, simple to understand. An abductive inference proceeds by "proposing" or generating hypotheses which, together with certain domain information, would account for or "explain" a group of facts or observations. For example[1], suppose we know that when it rains the grass gets wet. One day we observe that the grass is wet and hypothesise that it may have rained. Of course, there may be other reasons for the grass being wet — someone may have turned the sprinkler on for instance. There may be many possible hypotheses that would account for the new data — some mutually consistent, others conflicting. Often, though, we are interested in determining a "best" reason for our observations.

The description of abduction in the previous paragraph is still rather vague — terms such as "account" and "explain" are not very precise. The main aim of this paper is to present a more formal definition of abduction and to investigate certain different types of abduction within a logical setting. Our motivation stems from the lack of precise formal definitions and investigations into various types of abduction to be found in the literature. (Occasionally, a discussion of some aspects of abduction may be found in the literature but only in terms of a more restrictive language.)

Our main concern here is with the logical aspects of abduction though we briefly mention some non-logical considerations. In particular, we shall phrase our definitions in terms of a consequence relation $\vdash$. The advantage of this approach is that we can alter the abduction performed by changing the underlying logic provided its consequence relation still satisfies certain assumed properties. In the following section we present the assumptions we are making with regard to the language and the underlying logic. Section

[1] This example is based on an example presented in [9].

3 presents a definition of abduction and investigates a number of different types of abductions providing each with a definition based on logical criteria. A brief discussion of some of the special situations that may arise is presented in section 4. Our conclusions and suggestions for future work are given in section 5.

2 Preliminaries

We consider a fixed language $\mathcal{L}$ which consists of the standard logical connectives $\neg$, $\wedge$, $\vee$, $\rightarrow$, $\leftrightarrow$ and the symbol $\perp$ (falsum). Formulae in $\mathcal{L}$ will be denoted by lower case Greek characters $\phi, \psi, \ldots$ and sets of formulae will be denoted by upper case Greek characters $\Gamma, \Delta, \ldots$. The underlying logic will be identified with its consequence relation $\vdash \subseteq 2^{\mathcal{L}} \times \mathcal{L}$ which is assumed to possess the following properties:

If α is a truth-functional tautology, then $\vdash \alpha$	(Supraclassicality)
If $\{\alpha\} \vdash \beta$ and $\vdash \alpha$, then $\vdash \beta$	(Modus Ponens)
$\vdash$ is consistent (i.e., $\nvdash \perp$)	(Consistency)
$\Gamma \cup \{\alpha\} \vdash \beta$ iff $\Gamma \vdash \alpha \rightarrow \beta$	(Deduction Theorem)
If $\Gamma \vdash \alpha$ then $\Gamma' \vdash \alpha$ for some finite $\Gamma' \subseteq \Gamma$	(Compactness)

Note that, from Deduction and Compactness, it follows that $\vdash$ is monotonic (i.e., if $\Gamma \vdash \phi$ then $\Gamma \cup \{\psi\} \vdash \phi$). We make no further assumptions regarding the nature of $\mathcal{L}$ or the underlying logic.

3 Abduction

We begin by presenting a more precise definition of abduction than that outlined in the introduction and then proceed to define a number of different types of abduction.

3.1 Defining Abduction

The definition of abduction that we shall begin with is a fairly standard one, commonly found in abduction literature (see, for example, [8] [18]).

Definition 3.1 *An abduction*[2] *of a formula ϕ with respect to a domain theory Γ (a set of formulae) is a set of formulae E such that:*

[2] This use of the term abduction, to refer to both the particular process of inference and to the result of that inference, may prove a little disconcerting to some readers. However, we note that similar definitions can be found, in the literature, for deduction ([4] pp. 102-104, for instance).

(i) $\Gamma \cup E \vdash \phi$;

(ii) $\Gamma \cup E$ *is consistent (i.e.,* $\Gamma \cup E \nvdash \bot$*).*

An abduction is finite if the set E contains finitely many formulae.

For example, suppose $\Gamma = \{rain \rightarrow wet\_grass,\ sprinkler \rightarrow wet\_grass,\ storm\_clouds \wedge cold \rightarrow rain\ \}$ and $\phi = wet\_grass$ then possible abductions include $\{rain\}$, $\{sprinkler\}$, $\{rain \wedge sprinkler\}$, $\{rain, sprinkler\}$ and $\{storm\_clouds \wedge cold\}$.

Unlike our definition, some authors ([8], for example) consider abductions of a set of formulae Δ instead of a single formula ϕ. We shall concern ourselves with abductions of a single formula here as it would seem to be a more fundamental consideration even though the two approaches are quite similar.

With the above definition we can make the following observation which is an immediate consequence of the compactness of $\vdash$.

Observation 3.1 *If an abduction E of a formula ϕ with respect to a domain theory Γ exists then a finite abduction $E' \subseteq E$ (where E' is understood to be finite) of ϕ with respect to Γ exists.*

The following result concerning finite abductions is also easily proved.

Lemma 3.2 *If a finite abduction E of ϕ with respect to Γ exists then it can be represented by a single formula ψ (i.e.,* $\Gamma \cup \{\psi\} \vdash \phi$, $\Gamma \cup \{\psi\} \nvdash \bot$*).*

We can actually strengthen this result as it also holds for finitely axiomatisable abductions E.

This motivates the following rephrasing of our definition of abduction.

Definition 3.2 *An abduction of a formula ϕ with respect to a domain theory Γ is a formula ψ such that:*

(i) $\Gamma \cup \{\psi\} \vdash \phi$;

(ii) $\Gamma \cup \{\psi\}$ *is consistent (i.e.,* $\Gamma \cup \{\psi\} \nvdash \bot$*).*

We also say that ψ is abduced from Γ and ϕ.

In general this definition of abduction will permit a number of abductions for a given observation ϕ. This characterisation of abduction will prove helpful in providing relatively simple definitions for particular types of abduction. This also makes explicit the fact that we are only interested in dealing with abductions which are finite in nature.

3.2 Minimality

One consideration that is often deemed significant when determining abductions is to assume as little as possible in accounting for a formula ϕ. A reason for this is that we may wish to avoid superfluous abductions (i.e., abductions that may be irrelevant to the formula ϕ). A first attempt at accomplishing this within a logical setting would be to use the consequence relation $\vdash$ to define a notion of one abduction being "weaker" than another.

Definition 3.3 *An abduction ψ of ϕ with respect to Γ is weaker than an abduction ψ' of ϕ with respect to Γ if and only if $\psi' \vdash \psi$. We write this as $\psi \leq_{\Gamma,\phi} \psi'$.*
An abduction ψ of ϕ with respect to Γ is minimal iff $\psi \leq_{\Gamma,\phi} \psi'$ for all other abductions ψ' of ϕ with respect to Γ.

The fact that it has rained is a possible abduction for the grass being wet; so is the fact that the sprinkler was turned on. Of course, we may not be able to assume either of these with any great certainty and so we may be better off assuming that either it rained or the sprinkler was turned on (i.e., $rain \vee sprinkler$ is a possible abduction and $rain \vee sprinkler \leq_{\Gamma,\phi} rain$, $rain \vee sprinkler \leq_{\Gamma,\phi} sprinkler$).

Observation 3.3 *The "weakness" relation $\leq_{\Gamma,\phi}$ induces a partial order over the set of abductions of ϕ with respect to Γ.*

It may in fact be argued that this result should be weakened to state that $\leq_{\Gamma,\phi}$ is a partial pre-order. In determining whether the ordering $\leq_{\Gamma,\phi}$ satisfies the property of anti-symmetry we have taken the identity $=$ to mean logical equivalence (i.e., $\vdash \phi \leftrightarrow \psi$) rather than syntactic equivalence of the two formulae in question when $\phi \leq_{\Gamma,\phi} \psi$ and $\psi \leq_{\Gamma,\phi} \phi$. Effectively, we are treating formulae as propositions rather than sentences. The rationale for this is that the abductions in turn become the objects of the agent's beliefs and the objects of beliefs are normally taken to be propositions. We find a similar situation in the belief revision literature ([6] p. 55) where we may be more interested in the content of the data rather its particular formulation.

Intuitively, a "minimal" abduction under this weakness ordering would be some form of "maximal" disjunction of possible abductions. This, unfortunately, may not be expressible by any finite formula in the language!

Observation 3.4 *If there is an expressible minimal abduction ψ of ϕ with respect to Γ under $\leq_{\Gamma,\phi}$ then $\psi \leq_{\Gamma,\phi} \phi$.*

So, if a minimal abduction exists, it must be weaker than the new data ϕ under the ordering given by $\leq_{\Gamma,\phi}$.

Theorem 3.5 *For any abduction ψ of ϕ with respect to Γ weaker than ϕ, $\Gamma \cup \{\psi\} \vdash \delta$ iff $\Gamma \cup \{\phi\} \vdash \delta$ for any $\delta \in \mathcal{L}$.*

This result has important consequences for (abductive) system dynamics. So far we have assumed a fixed domain theory Γ, and have, as it were, confined ourselves to *static* abduction. However any interesting theory of abduction must allow for *dynamic abduction*. An obvious way to deal with this is to let the chosen abduction ψ in light of observation ϕ help the domain theory Γ evolve to $\Gamma^{@}_{\psi}$. The nature of $\Gamma^{@}_{\psi}$ will be partially determined by what system constraints are allowed to operate. The only relevant system constraint that we have explicitly allowed so far is that every domain theory must be a set of sentences or propositions.

It is reasonable to allow the following idealisation:

- **System Closure:** every domain theory Γ is a closed theory, i.e., if $\Gamma \vdash \alpha$ then $\alpha \in \Gamma$.

A virtue of this constraint is that it vastly simplifies the account of belief updating (cf. [1]). Given this constraint, $\Gamma^{@}_{\psi}$ may be identified with the deductive closure of $\Gamma \cup \{\psi\}$. From Theorem 3.5, it then follows that for every abduction ψ weaker than ϕ, the new domain theory $\Gamma^{@}_{\psi}$ is just the deductive closure of $\Gamma \cup \{\phi\}$. Thus the difference among various abductions ψ weaker than ϕ is effectively obliterated. This provides a reason for not requiring the system closure constraint to be satisfied; that is, for allowing the background theory to be represented as an open theory. We find an analogous situation in the literature on belief change — many writers have advocated the use of belief bases instead of belief sets in offering a reasoned account of belief revision [14] [13]. If we allow a finite representation of the domain theory Γ, then the set $\Gamma^{@}_{\psi}$ can simply be identified with $\Gamma \cup \{\psi\}$. It must, however, be noted that as long as ψ is weaker than ϕ, the deductive closure of $\Gamma^{@}_{\psi}$ will always suffer from the problem outlined above.

On the other hand, instead of fixing this problem by requiring the domain theory to be finitely represented, one might question the rationality of abductions weaker than the observation ϕ itself. Intuitively, an abduction weaker than the observation we are trying to account for would not seem to make much sense. If we observe that the grass is wet, then assuming (or abducing) that either the grass is wet or it rained would not seem to be very sensible. These observations suggest that this sort of abduction (and minimal abductions under this ordering, in particular, if they exist) is best avoided. In the next section this type of abduction is used to characterise a notion of triviality akin to Jackson's [8].

However, there is another notion of minimality, is proposed by Jackson [8] which does not suffer from the above shortcomings. Jackson adopts a propositional calculus with formulae in clausal form and phrases his definition in terms of set theoretic properties.

Definition 3.4 *(Jackson-minimality[3])*
An explanation of a data set Δ in terms of causal axioms Γ is a set of clauses E such that $\Gamma, E \vdash \Delta$ and $\Gamma \cup E$ is consistent. E is minimal iff $\Gamma, E' \not\vdash \Delta$ for any $E' \subset E$.

The resolution-based procedure given for determining abductions (called explanations by Jackson) only produces abductions which are conjunctions of literals, contrary to the indications of this definition. The minimal abductions in this sense are those conjunctions of literals, for which there is no other conjunct, containing a proper subset of these literals, which is also an abduction. This type of abduction is also advocated by Reiter and De Kleer [21] and is referred to as the "simplest explanation" ([21] p. 90). We shall now show how to characterise such abductions within our framework.

Consider an abduction expressed in disjunctive normal form

$$\psi \equiv (l_{11} \wedge \ldots \wedge l_{1n_1}) \vee \ldots \vee (l_{m1} \wedge \ldots \wedge l_{1n_m})$$

Any of the (conjunctive) disjuncts $(l_{i1} \wedge \ldots \wedge l_{in_i})$ would suffice as an abduction. Moreover, as we have seen above, certain disjunctive abductions may not be very useful. Here, we are only interested in considering conjunctive abductions in order to mimic Jackson's definition so, we shall call an abduction *conjunctive minimal* if its disjunctive normal form consists of a single disjunct and there is no other abduction whose disjunctive normal

[3] We paraphrase the definition in Jackson [8].

form (DNF) consists of a single disjunct containing a proper subset of its literals. Formally:

Definition 3.5 *An abduction ψ of ϕ with respect to Γ is conjunctive minimal iff $DNF(\psi) = C_{\psi}$ is a single disjunct and there is no other abduction ψ' of ϕ with respect to Γ where $DNF(\psi') = C_{\psi'}$ is a single disjunct for which $\{C_{\psi}\} \vdash C_{\psi'}$ and $\{C_{\psi'}\} \not\vdash C_{\psi}$.*

In other words, we are looking for the weakest abductions whose DNF is a single disjunct (i.e., merely a conjunction of literals). Such "conjunctive abductions" are very appealing in terms of syntactic simplicity and "ease" of calculation. However, intuitively, abductions containing other connectives would appear to make more sense in certain situations.

There are other notions of minimality (some also called *simplicity*) criteria in the literature but not all of these are based on logical criteria ([10] for instance).

3.3 Triviality

We now discuss Jackson's [8] notion of triviality within our framework. From our definition of abduction it should be fairly clear that if we wish to obtain an abduction of a formula ϕ with respect to any domain theory Γ then we can simply take the abduction to be ϕ itself (since $\{\phi\} \vdash \phi$ and therefore $\Gamma \cup \{\phi\} \vdash \phi$) if ϕ does not contradict Γ. Of course, in doing this we are gaining no new knowledge about our domain theory apart from ϕ and its deductive consequences. Moreover, by Theorem 3.5, abductions weaker than ϕ do not help us to gain any new knowledge either. So, we may wish to disregard such abductions.

Within our framework we can express this in the following way:

Definition 3.6 *(non-trivial abduction)*
An abduction ψ of ϕ with respect to Γ is trivial iff $\psi \leq_{\Gamma,\phi} \phi$. If an abduction is not trivial then it is non-trivial.

In the example introduced in section 3.1 simply hypothesising that the grass is wet would account for the observations. So, wet_grass is a trivial abduction while $rain$, $sprinkler$ and $sprinkler \wedge rain$ are non-trivial.

Since $\leq_{\Gamma,\phi}$ is a partial ordering, the only abductions ψ of ϕ with respect to Γ weaker than ϕ (i.e, $\psi \leq_{\Gamma,\phi} \phi$) such that $\phi \leq_{\Gamma,\phi} \psi$ are those such that $\vdash \phi \leftrightarrow \psi$. We call these *most trivial abductions* of ϕ with respect to Γ.

3.4 Specificity

Stickel [23] introduces the notions of *most specific abduction* and *least specific abduction*. These are described within a resolution-based system for abductive reasoning. A most specific abduction allows only pure literals (literals that resolve with no other clause in a knowledge base) to be used in an abduction. If implications in a domain theory Γ are considered to represent causal relationships then this type of abduction would correspond to "backward chaining" along the causal chains as far as possible. Stickel claims that this type of abduction is suited to diagnostic tasks. This type of abduction is also adopted in [19]. Least specific abduction on the other hand takes the most trivial abduction in the sense of the previous section. Stickel argues that this form of abduction is preferable in natural-language interpretation.

Stickel, however, does not provide a formal definition of these concepts. We shall introduce three different types of specificity. In some cases we shall proceed to examine the most and least specific abductions produced. Furthermore, we can generalise these two extremes to gain a number of *levels of specificity*.

The first type of specificity we shall investigate considers one abduction to be more specific than another if it can be used togther with the domain theory to prove this second abduction. This means that it was somehow further "back" along an implication "chain" and therefore likely to add more information than the other abduction. We can express this in the following manner.

Definition 3.7 *(relative specificity)*
An abduction ψ of ϕ with respect to Γ is relatively more specific than an abduction ψ' of ϕ with respect to Γ iff $\Gamma \cup \{\psi\} \vdash \psi'$. We write $\psi' \preceq_{\Gamma,\phi} \psi$. We also say that ψ is a relatively least (respectively most) specific abduction of ϕ with respect to Γ iff $\psi \preceq_{\Gamma,\phi} \psi'$ ($\psi' \preceq_{\Gamma,\phi} \psi$) for all abductions ψ' of ϕ with respect to Γ such that $\not\vdash \psi \leftrightarrow \psi'$.

In the example that we have been using thus far, we have the following relative specificities:

$$wet\_grass \preceq_{\Gamma,\phi} rain \preceq_{\Gamma,\phi} rain \wedge wet\_grass \preceq_{\Gamma,\phi} storm\_clouds \wedge cold$$

$$wet\_grass \preceq_{\Gamma,\phi} sprinkler \preceq_{\Gamma,\phi} sprinkler \wedge rain$$

The ordering generated by such a definition may be characterised as follows.

Observation 3.6 *The relatively more specific relation $\preceq_{\Gamma,\phi}$ is a partial preorder over the set of abductions of ϕ with respect to Γ.*

Now, since $\Gamma \cup \{\psi\} \vdash \phi$ for any abduction ψ of ϕ, then every abduction of ϕ is more specific than ϕ (i.e., $\phi \preceq_{\Gamma,\phi} \psi$) and therefore ϕ is a least element in this ordering. Moreover, other possible least specific abductions are characterised by the following result.

Lemma 3.7 *An abduction ψ of ϕ with respect to Γ is a relatively least specific abduction of ϕ with respect to Γ iff $\Gamma \vdash \phi \leftrightarrow \psi$.*

This result allows us to extend our previous observation by noting that the relatively more specific ordering is in fact a well-founded ordering. In fact, this ordering can be utilised as a way of arranging abductions into "levels of specificity". The relatively least and most (if any exist) specific abductions can be considered at opposite ends of this "spectrum" of levels. Maximal length chains of distinct formulae under the above ordering can then be used to determine a particular formula's level.

One problem with this definition is that it includes abductions that prove other abductions without the aid of the domain theory ($\{\psi\} \vdash \psi'$ and therefore we have $\Gamma \cup \{\psi\} \vdash \psi'$) in the ordering while we may be interested solely in abductions which establish a "specificity sequence" with the aid of the domain theory. Reasons for disregarding such abductions will be addressed in Section 4.2 for abductions that prove the new data ϕ without the help of the domain theory but these arguments can also be applied to the above case. To handle this situation we introduce the notion of strong relative specificity.

Definition 3.8 *(strong relative specificity)*
An abduction ψ of ϕ with respect to Γ is strongly relatively more specific than an abduction ψ' of ϕ with respect to Γ iff ψ is relatively more specific than ψ' but $\psi \nvdash \psi'$.

Unfortunately, the resulting relation does not even satisfy transitivity (though it is irreflexive). It is therefore difficult to see how to sensibly produce levels of specificity or least and most elements.

For completeness we also introduce the following type of specificity.

Definition 3.9 *(absolute specificity)*
An abduction ψ of ϕ with respect to Γ is absolutely more specific than an abduction ψ' of ϕ with respect to Γ iff $\{\psi\} \vdash \psi'$.

Notice that this ordering (and, in fact, the definition) is exactly the same as the weakness ordering introduced in Section 3.2 and therefore possesses the same properties. In particular, it is a partial order, if we consider the logical content of the formulae, which may not have an expressible minimal element.

3.5 Selection

We noted in the introduction that abduction is often referred to as "inference to the best explanation". We have already seen a number of different types of abduction and supplied each with a definition based solely on logical criteria. Each of these different forms of abduction is capable, in some way, of reducing the number of possible abductions to consider but they may not provide us with a single "best" abduction.

It is our belief that logical criteria alone may not be capable of singling out such an abduction. One way to determine a single "best" abduction is to adopt some form of non-logical criteria to further select between the remaining abductions singled out by one or more of the logical criteria described above. Since, in this paper, we are interested only in investigating the logical aspects of abductive reasoning as we have done thus far, we shall not study any notions of further selecting between abductions. However, to give the reader an idea of how this may be done, we shall mention a few of the techniques found in the literature.

Stickel [23] adopts a cost-based measure; Peng and Reggia [17] develop a probabilistic approach; Charniak [2] explores simplicity measures; coherence based approaches have been investigated by Thagard [24] and Ng and Mooney [15]; and, utility-based measures have been suggested by Ram and Leake [20].

4 Discussion

Thus far we have considered abduction to be the process whereby we find some formula ψ which, when added to our domain theory Γ, allows us to prove the new data ϕ. Some readers may be troubled by the fact that it appears that we are implicitly assuming:

- the domain theory does not entail our new data ($\Gamma \not\vdash \phi$)
- an abduction ψ cannot prove the new data ϕ without the help of the domain theory ($\{\psi\} \not\vdash \phi$).

We shall now investigate what occurs if these conditions are not satisfied.

4.1 Domain Theory Entails the Observation

In the case where the domain theory Γ already accounts for the new data ϕ an interesting situation occurs.

Observation 4.1 *If $\Gamma \vdash \phi$ then any $\delta \in \mathcal{L}$ consistent with Γ (i.e., $\Gamma \cup \{\delta\} \nvdash \perp$) is an abduction of ϕ with respect to Γ.*

So, any formula consistent with the domain theory Γ would be a possible abduction. This is certainly undesirable.

In such a case one may be tempted to suggest that no abduction is necessary or proceed and in fact consider any formula consistent with Γ but, in many cases, we may be interested in determining a more "relevant" abduction. One way to achieve this is to consider some subset of the domain theory which does not prove the new data. This motivates the following revision of our definition of abduction.

Definition 4.1 *An abduction of a formula ϕ with respect to a domain theory Γ is a formula ψ such that the following conditions are satisfied:*

(i) *take some $\Gamma' \subseteq \Gamma$ such that $\Gamma' \nvdash \phi$*

(ii) $\Gamma' \cup \{\psi\} \vdash \phi$

(iii) $\Gamma \cup \{\psi\} \nvdash \perp$

Notice that this definition does not deal with the problem of determining the nature of Γ', the subset of the domain theory to be used in determining an abduction. Many possible Γ's may exist together with many methods for producing such a subset. We shall not provide any specific criteria here but, rather, direct the interested reader to the literature on the process of belief contraction [1] [6] which may be used for this purpose.

Note also that there are no further restrictions placed on the derived abduction ψ. In particular, it may actually result from that part removed from our original domain theory (i.e., $\Gamma - \Gamma' \vdash \psi$) in which case the "best" abduction was in fact contained in our original domain theory. This may be viewed as additional support for the utility of ψ. This type of abduction, if preferred, can also be viewed as an attempt by the agent to perform introspection.

Some may suggest that a similar analysis could be applied to the situation in which the negation of the new data is a consequence of the domain theory (i.e., $\Gamma \vdash \neg\phi$). We shall not pursue this here but suggest that no abduction exists in this case[4].

4.2 Abduction Proves Observation

The other interesting case occurs when the potential abduction ψ of ϕ proves the new data ($\{\psi\} \vdash \phi$ — provided, of course, that ψ is consistent with the domain theory Γ) for in this case $\Gamma \cup \{\psi\} \vdash \phi$ trivially. In this case we can consider the abduction to occur with respect to a theorem of the logic (viz. $\vdash \psi \rightarrow \phi$ — via the Deduction Theorem). In other words, the domain theory provides no "help" in determining the abduction. Notice that these types of abduction are inherent in the logic in a certain sense and therefore may always be performed regardless of the contents of the domain theory (only consistency with the domain theory is required).

Interestingly, there are similarities between this type of abduction and some of what is termed induction in the machine learning literature. The inverse resolution [12] W-operators of inter and intra-construction for example. If we have an (incomplete) domain theory Γ and are presented with an example ϕ which is not accounted for by Γ (i.e., $\Gamma \not\vdash \phi$) then inverse resolution provides a number of operators which attempt to construct a new domain theory Γ' such that $\Gamma' \vdash \Gamma$ and $\Gamma' \vdash \phi$[5]. These operators work by attempting to invert the resolution process.

The W-operators determine a new domain theory Γ' by introducing terms not present in the original domain theory. The propositional inter-construction operator for instance may be represented by the following schema

$$\frac{\alpha \wedge \beta \rightarrow X, \alpha \wedge \gamma \rightarrow Y}{\beta \wedge Z \rightarrow X, \gamma \wedge Z \rightarrow Y, \alpha \rightarrow Z}$$

(where α, β, γ are conjunctions of literals and X, Y, Z represent single literals). In this case, the top line represents new information (one of the formulae may in fact come from the original domain theory) while the bottom line represents new formulae to be added to the original domain theory. We can represent formulae matching this schema as theorems in our logic $\vdash [(\beta \wedge Z \rightarrow X) \wedge (\gamma \wedge Z \rightarrow Y) \wedge (\alpha \rightarrow Z)] \rightarrow [(\alpha \wedge \beta \rightarrow X) \wedge (\alpha \wedge \gamma \rightarrow Y)]$ In this case, an abduction for $\phi = (\alpha \wedge \beta \rightarrow X) \wedge (\alpha \wedge \gamma \rightarrow Y)$ is $\psi = (\beta \wedge Z \rightarrow X) \wedge (\gamma \wedge Z \rightarrow$

[4] To see how this may be handled in the context of abductive belief revision, consult [16].

[5] Note the similarity to the dynamic abductive scenario presented in Section 3.2.

$Y) \wedge (\alpha \rightarrow Z)$ where $\psi \vdash \phi$. This however, does not hold for the V-operators of absorption and identification. In fact, the above observation can be seen as a manifestation of the observation by Console et. al. [3] that the W-operators can be considered to use an empty theory ($\Gamma = \emptyset$).

This type of induction is not new. Morgan [11] presents a method for performing induction by inverting the inference rules of a logical system. It is in fact stated [22] that a formula ψ is more general than a formula ϕ if $\psi \vdash \phi$. It is interesting to note that we use this concept to define the notion of absolute specificity (and weakness).

In light of these observations, it may be tempting to posit that such inductive techniques can be thought of as special cases of abduction where we are "backward chaining" (moving from consequent to antecedent) over formulae that are theorems of the logic rather than formulae that are solely consequences of the domain theory (which may be considered a characteristic of pure abduction). Interestingly enough, Harman [7] also considers enumerative induction to be subsumed by the process of abduction[6].

We may therefore wish to constrain our previous definition even further in order to characterise "pure" abduction.

Definition 4.2 *An abduction of a formula ϕ with respect to a domain theory Γ is a formula ψ such that the following conditions are satisified:*

(i) $\Gamma' \subseteq \Gamma$ *such that* $\Gamma' \nvdash \phi$

(ii) $\Gamma' \cup \{\psi\} \vdash \phi$

(iii) $\psi \nvdash \phi$

(iv) $\Gamma \cup \{\psi\} \nvdash \perp$

5 Conclusions and Future Work

We have investigated a general logical framework for defining the notion of abduction. The notions of minimal, trivial and specific abductions were also investigated and defined in the same logical style together with an analysis of some resulting properties. The proposed definitions were phrased in terms of a consequence relation $\vdash$ based upon supraclassical propositional logic.

[6] Which he refers to as "inference to the best explanation".

Therefore, these results are also applicable to other logics provided they satisfy the requirements for such a logic.

Interesting situations occur if the domain theory already proves the new data or the potential abduction proves the new data without the assistance of the domain theory. In the first instance, any formula consistent with the domain theory is a possible abduction. A remedy for this was presented. In the second case, parallels were drawn with the process of induction. These considerations led to a revised definition of abduction. Moreover, together with the presentation of abduction, they exhaust all possible cases for $\Gamma \cup \{\psi\} \vdash \phi$.

Possibilities for future work include the investigation and definition of other possible types of abduction. The notion of selection was not studied here but a mechanism (most likely non-logical in nature) that integrates nicely with the definitions presented would be advantageous. Finally, having a uniform set of definitions of types of abduction and facts about their properties, it would be interesting to see how applicable they are to other areas of artificial intelligence (belief revision and scientific theory formation for instance).

References

[1] Carlos E. Alchourrón, Peter Gärdenfors, and David Makinson. On the logic of theory change: Partial meet contraction and revision functions. *Journal of Symbolic Logic*, 50:510–530, 1985.

[2] Eugene Charniak and Drew McDermott. *Introduction to Artificial Intelligence*. Addison-Wesley, 1985.

[3] Luca Console, Attilio Giordano, and Lorenza Saitta. Investigating the relationships between abduction and inverse resolution in propositional calculus. In Z W Ras and M Zemankova, editors, *Methodologies for Intelligent Systems : Proceedings of the 6th International Symposium (ISMIS '91)*, pages 316–325, 1991.

[4] Herbert B. Enderton. *Mathematical Introduction to Logic*. Academic Press, 1972.

[5] K. T. Fann. *Peirce's Theory of Abduction.* Martinus Nijhoff, The Hague, Holland, 1970.

[6] Peter Gärdenfors. *Knowledge in Flux: Modeling the Dynamics of Epistemic States.* Bradford Books, MIT Press, Cambridge Massachusetts, 1988.

[7] Gilbert H Harman. Inference to the best explanation. *Philosophical Review*, 74:88–95, January 1965.

[8] Peter Jackson. Propositional abductive logic. In *Proceedings of the Seventh Conference of the Society for the Study of Artificial Intelligence and Simulation of Behaviour*, pages 89–94, 1989.

[9] Kurt Konolige. A general theory of abduction. In Paul O'Rorke, editor, *Working Notes of the 1990 Spring Symposium on Automated Abduction*, pages 62–66. Technical Report 90-32, University of California, Irvine, 1990.

[10] Hector J. Levesque. A knowledge-level account of abduction. In *Proceedings of the 11th International Joint Conference on Artificial Intelligence*, pages 1061–1067. Morgan Kaufman, 1989.

[11] Charles G. Morgan. Hypothesis generation by machine. *Artificial Intelligence*, 2:179–187, 1971.

[12] Stephen Muggleton. Duce, an oracle based approach to constructive induction. In *Proceedings of the 10th International Joint Conference on Artificial Intelligence*, pages 287–292. Morgan Kaufman, 1987.

[13] Abhaya C. Nayak. Foundational belief change. *Journal of Philosophical Logic*, 1994 (Forthcoming).

[14] Bernhard Nebel. Belief revision and default reasoning: Syntax based approaches. In Peter Gärdenfors, editor, *Belief Revision*, pages 52–88. Cambridge University Press, 1992.

[15] Hwee Tou Ng and Raymond J. Mooney. On the role of coherence in abductive explanation. In *Proceedings of the National Conference in Artificial Intelligence*, pages 337–342, 1990.

[16] Maurice Pagnucco, Abhaya Nayak, and Norman Y. Foo. Abductive expansion: The application of abductive inference to the process of belief change. submitted.

[17] Yun Peng and James A. Reggia. *Abductive Inference Models for Diagnostic Problem-Solving*. Springer-Verlag, 1990.

[18] David Poole. A logical framework for default reasoning. *Artificial Intelligence*, 36:27–47, 1988.

[19] H. E. Pople. On the mechanisation of abductive logic. In *Proceedings of the Third International Joint Conference on Artificial Intelligence*, pages 147–152, 1973.

[20] Ashwin Ram and David Leake. Evaluation of explanatory hypotheses. In *Proceedings of the Thirteenth Annual Conference of the Cognitive Science Society*, pages 867–871, 1991.

[21] Raymond Reiter and Johan de Kleer. Foundations of assumption-based truth maintence systems: Preliminary report. In *Proceedings of the National Conference in Artificial Intelligence*, pages 183–188, 1987.

[22] Celine Rouveirol and Jean Francois Puget. Beyond inversion of resolution. In *Proceedings of the 7th International Machine Learning Workshop*, pages 122–130. Morgan Kaufman, 1990.

[23] Mark E. Stickel. A prolog-like inference system for computing minimal-cost abductive explanations in natural-language interpretation. *Annals of Mathematics and Artificial Intelligence*, 4:89–106, 1991.

[24] Paul Thagard. Explanatory coherence. *Behavioral and Brain Sciences*, 12:435–502, 1989.