Determining Explanations using Knowledge Transmutations

M.A. Williams

maryanne@frey.newcastle.edu.au

Information Systems Group
Department of Management
University of Newcastle, NSW 2308
Australia

Knowledge Systems Group
Basser Department of Computer Science
University of Sydney, NSW 2006
Australia

Telephone: +61 49 215996 Facsimile: +61 49 216911

M. Pagnucco

morri@cs.su.oz.au
Knowledge Systems Group
Basser Department of Computer Science
Madsen Building, F09
University of Sydney, NSW 2006
Australia
Telephone: +61 2 6923423 Facsimile: +61 2 6923838

N.Y. Foo

norman@cs.su.oz.au
Knowledge Systems Group
Basser Department of Computer Science
Madsen Building, F09
University of Sydney, NSW 2006
Australia
Telephone: +61 2 6923423 Facsimile: +61 2 6923838

B. Sims

bsims@frey.newcastle.edu.au
Department of Mathematics
University of Newcastle, NSW 2308
Australia
Telephone: +61 49 215526 Facsimile: +61 49 216900

Keywords: belief management and revision, abduction, logics of knowledge and belief, explanation.

1 Introduction

Knowledge systems represent a view of the world. That is, an incomplete description of the world state. Consequently they require a mechanism for modifying their view as more information about the world is acquired.

The AGM paradigm is a formal approach to informational change. Moreover, it provides a mechanism for revision and contraction of information. Within the AGM paradigm [1,2,4] the family of revision functions, and the family of contraction functions are constrained by rationality postulates. The logical properties of a knowledge system are not strong enough to uniquely determine a revision or contraction function, therefore the principal constructions for these functions rely on some form of underlying preference relation, such as a family of selection functions [1,2], a system of spheres [5], a nice preorder on models [8], or an epistemic entrenchment ordering [4]. In this light we consider a knowledge system to be composed of a set of knowledge together with an underlying preference relation.

We refer to the process of changing the underlying preference relation for a knowledge system as a transmutation. In [15] Williams described transmutations which essentially generalize the work of Spohn [13]. Spohn argues that *conditionalization* [13] which is a constructive method for revising or contracting a knowledge system using ordinal conditional functions, is an appropriate transmutation. Williams [15] explored conditionalization further and examined an alternative construction; an *adjustment*. Intuitively adjustments involve the 'minimal change' to an ordinal conditional function necessary to effect the revision or contraction.

Spohn [12] characterizes the notion of "reason for", and shows how it can be captured within his framework in [12, 13].

The objective of this paper is firstly, to recast Spohn's notion of "reason for" in a more general setting using a transmutation rather than a conditionalization, and secondly, to use this approach to determine a most plausible explanation(s). We show that an adjustment determines intuitively satisfying results, while conditionalization does not. Thirdly, we explore the relationship between explanation based on adjustments and explanation based on abduction.

For completeness and to set the notation we briefly outline the AGM paradigm in section 2. In section 3, we describe ordinal epistemic entrenchment functions, and in section 4 we describe their transmutations. In particular, we describe two types of transmutations; conditionalizations and adjustments. In section 5, we outline Spohn's notion of explanation, and provide conditions which allow this notion to be defined using arbitary transmutations, conditionalizations, and adjustments. In section 6 we illustrate these definitions and results with examples. In section 7 we compare our notion of explanation with one based upon the process of abductive inference.

We begin with some technical preliminaries. Let $\mathbf{L}$ denote a language which contains a complete set of Boolean connectives. We will denote sentences in $\mathbf{L}$ by lower case Roman letters. We assume $\mathbf{L}$ is governed by a logic that is identified with its consequence relation

$\vdash$. The relation $\vdash$ is assumed to satisfy the following conditions:

(a) If a is a truth-functional tautology, then $\vdash \mathsf{a}$.

(b) If $\mathsf{a} \vdash \mathsf{b}$ and $\vdash \mathsf{a}$, then $\vdash \mathsf{b}$ (*modus ponens*).

(c) $\vdash$ is consistent, that is, $\nvdash \bot$, where $\bot$ denotes the inconsistent theory.

(d) $\vdash$ satisfies the deduction theorem.

(e) $\vdash$ is compact.

The set of all logical consequences of a set $T \subseteq \mathbf{L}$, that is $\{\mathsf{a} : T \vdash \mathsf{a}\}$, is denoted by $\mathrm{Cn}(T)$. The set of tautologies, $\{\mathsf{a} : \vdash \mathsf{a}\}$, is denoted by $\top$, and those formulae not in $\top$ are referred to as *nontautological*. A *theory* of $\mathbf{L}$ is any subset of $\mathbf{L}$, closed under $\vdash$. A *consistent* theory of $\mathbf{L}$ is any theory of $\mathbf{L}$ that does not contain both a and $\neg\mathsf{a}$, for any formula a of $\mathbf{L}$. A theory is *finite* if there is a finite number of equivalent formulas relative to $\vdash$. We denote the set of all theories of $\mathbf{L}$ by $\mathcal{K}_{\mathbf{L}}$. The *dual atoms* for a finite theory T are those nontautological elements $\mathsf{a} \in T$ such that for all $\mathsf{b} \in \mathrm{Cn}(\{\mathsf{a}\}) \cap T$, either $\vdash \mathsf{b}$ or $\vdash \mathsf{a} \equiv \mathsf{b}$ [11], that is, the maximal disjunctions of atoms.

2 The AGM Paradigm

In the AGM paradigm knowledge sets are taken to be theories, and informational changes are therefore regarded as transformations on theories. There are three principal types of AGM transformations; expansion, contraction and revision. These transformations allow us to model changes of information based on the *principle of minimal change*. Expansion is the simplest change, it models the incorporation of a formula that is consistent with the knowledge set. The *expansion* of a theory T with respect to a sentence a, denoted as T^+_{a}, is defined to be the logical closure of T and a, that is $T^+_{\mathsf{a}} = \mathrm{Cn}(T \cup \{\mathsf{a}\})$.

A *contraction* of T with respect to a, T^-_{a}, involves the removal of a set of sentences from T so that a is no longer implied. Formally, a well-behaved contraction operator $^-$ is any function from $\mathcal{K}_{\mathbf{L}} \times \mathbf{L}$ to $\mathcal{K}_{\mathbf{L}}$, mapping (T, a) to T^-_{a} which satisfies the following postulates:

($^-$1) For any $\mathsf{a} \in \mathbf{L}$ and any $T \in \mathcal{K}_{\mathbf{L}}$, $T^-_{\mathsf{a}} \in \mathcal{K}_{\mathbf{L}}$

($^-$2) $T^-_{\mathsf{a}} \subseteq T$

($^-$3) If $\mathsf{a} \notin T$ then $T^-_{\mathsf{a}} = T$

($^-$4) If $\nvdash \mathsf{a}$ then $\mathsf{a} \notin T^-_{\mathsf{a}}$

($^-$5) $T \subseteq (T^-_{\mathsf{a}})^+_{\mathsf{a}}$

($^-$6) If $\vdash \mathsf{a} \equiv \mathsf{b}$ then $T^-_{\mathsf{a}} = T^-_{\mathsf{b}}$

($^-$7) $T^-_{\mathsf{a}} \cap T^-_{\mathsf{b}} \subseteq T^-_{\mathsf{a}\wedge\mathsf{b}}$

($^-$8) If $\mathsf{a} \notin T^-_{\mathsf{a}\wedge\mathsf{b}}$ then $T^-_{\mathsf{a}\wedge\mathsf{b}} \subseteq T^-_{\mathsf{a}}$

($^-$9) For every nonempty set Γ of nontautological formulae, there exists a formula $\mathsf{a} \in \Gamma$ such that $\mathsf{a} \notin T^-_{\mathsf{a}\wedge\mathsf{b}}$ for every $\mathsf{b} \in \Gamma$.

A *revision* attempts to transform a theory as "little as possible" in order to incorporate a sentence. Formally, a well-behaved† revision operator * is any function from $\mathcal{K}_{\mathbf{L}} \times \mathbf{L}$ to $\mathcal{K}_{\mathbf{L}}$, mapping (T, a) to T^*_{a} which satisfies the following postulates:

(*1) For any $\mathsf{a} \in \mathbf{L}$ and any $T \in \mathcal{K}_{\mathbf{L}}$, $T^*_{\mathsf{a}} \in \mathcal{K}_{\mathbf{L}}$
(*2) $\mathsf{a} \in T^*_{\mathsf{a}}$
(*3) $T^*_{\mathsf{a}} \subseteq T^+_{\mathsf{a}}$
(*4) If $\neg\mathsf{a} \notin T$ then $T^+_{\mathsf{a}} \subseteq T^*_{\mathsf{a}}$
(*5) $T^*_{\mathsf{a}} = \perp$ if and only if $\vdash \neg\mathsf{a}$
(*6) If $\vdash \mathsf{a} \equiv \mathsf{b}$ then $T^*_{\mathsf{a}} = T^*_{\mathsf{b}}$
(*7) $T^*_{\mathsf{a}\wedge\mathsf{b}} \subseteq (T^*_{\mathsf{a}})^+_{\mathsf{b}}$
(*8) If $\neg\mathsf{b} \notin T^*_{\mathsf{a}}$ then $(T^*_{\mathsf{a}})^+_{\mathsf{b}} \subseteq T^*_{\mathsf{a}\wedge\mathsf{b}}$
(*9) For every nonempty set Γ of nontautological formulae, there exists a formula $\mathsf{a} \in \Gamma$ such that $\mathsf{a} \notin T^*_{\neg\mathsf{a}\vee\neg\mathsf{b}}$ for every $\mathsf{b} \in \Gamma$.

Contraction and revision are interdefinable. In particular well-behaved contraction and well-behaved revision are related by means of the *Levi Identity*, $T^*_{\mathsf{a}} = (T^-_{\neg\mathsf{a}})^+_{\mathsf{a}}$, which defines a well-behaved revision in terms of a well-behaved contraction, and conversely by means of the *Harper Identity*, $T^-_{\mathsf{a}} = T \cap T^*_{\neg\mathsf{a}}$, which defines a well-behaved contraction in terms of a well-behaved revision.

3 Informational Change using Ordinal Functions

Spohn [13] introduced ordinal conditional functions, and in [15] Williams introduced a related structure, ordinal epistemic entrenchment functions, and established their connection. Essentially, ordinal conditional functions take possible worlds to ordinals, and ordinal epistemic entrenchment functions take formulae to ordinals. Intuitively, these functions define a ranking of the respective domains which provides a response schema for all "possible consistent information" [9].

In this section we briefly describe ordinal epistemic entrenchment functions. Formally we have the following definition.

`Definition`:

An *ordinal epistemic entrenchment function*, OEF, is a function $\mathbf{E}$ from the formulae in $\mathbf{L}$ into the class of ordinals such that the following conditions are satisfied.

(OEF1) For all a, $\mathsf{b} \in \mathbf{L}$, if $\mathsf{a} \vdash \mathsf{b}$, then $\mathbf{E}(\mathsf{a}) \leq \mathbf{E}(\mathsf{b})$
(OEF2) For all a, $\mathsf{b} \in \mathbf{L}$, $\mathbf{E}(\mathsf{a}) \leq \mathbf{E}(\mathsf{a} \wedge \mathsf{b})$ or $\mathbf{E}(\mathsf{b}) \leq \mathbf{E}(\mathsf{a} \wedge \mathsf{b})$
(OEF3) $\vdash \mathsf{a}$ if and only if $\mathbf{E}(\mathsf{a}) \geq \mathbf{E}(\mathsf{b})$ for all $\mathsf{b} \in \mathbf{L}$
(OEF4) If a is inconsistent, then $\mathbf{E}(\mathsf{a}) = 0$.

Intuitively, an OEF represents an *epistemic entrenchment* [4] grading of formulae, where the higher the ordinal assigned to a formula the more entrenched it is. We denote the family of all ordinal epistemic entrenchment functions by $\mathcal{E}$. In the definition below we identify the knowledge set represented by an OEF.

† (*9) is not identical with the well-behaved postulate in [5], however the same family of revision functions is obtained.

`Definition`:
We define the *knowledge set* represented by $\mathbf{E} \in \mathcal{E}$ to be $\{\mathsf{a} : \ \mathbf{E}(\mathsf{a}) > 0\}$, and denote it by $\mathrm{ks}(\mathbf{E})$.

We note that $\mathrm{ks}(\mathbf{E})$ is a theory. Gärdenfors and Makinson [4] show that for finite languages, an epistemic entrenchment ordering is determined by the dual atoms. Similarly, we can describe an ordinal epistemic entrenchment function for a finite language by assigning an ordinal to each dual atom. The ordinal assigned to each formula in $\mathbf{L}$ is then uniquely determined by (OEF1) – (OEF4).

`Definition`:
Given an $\mathbf{E} \in \mathcal{E}$. A formula a is *accepted* if and only if $\mathsf{a} \in \mathrm{ks}(\mathbf{E})$. If a and b are both accepted, then a is more *firmly accepted* than b if and only if $\mathbf{E}(\mathsf{a}) > \mathbf{E}(\mathsf{b})$. Moreover, we refer to $\mathbf{E}(\mathsf{a})$ as the *degree of acceptance* for the formula a. A formula a is more *plausible* than a formula b if and only if either $\mathbf{E}(\mathsf{a}) > \mathbf{E}(\mathsf{b})$ or $\mathbf{E}(\neg\mathsf{b}) > \mathbf{E}(\neg\mathsf{a})$.

Hence it can be seen that the knowledge set represented by an OEF, $\mathrm{ks}(\mathbf{E})$ is the set of formulae whose degree of acceptance is greater than zero.

4 Transmutations

In [15] Williams provided conditions which determined both well-behaved contraction and well-behaved revision operators from an OEF, and conversely, showed that for every well-behaved operator there exists an OEF satisfying these conditions.†

For contraction and revision the information inputs are a formula. We now define a transmutation of OEF's where the informational input is a formula a and an ordinal i. The interpretation [2] of this is that the formula a is the information to be accepted by the knowledge set, and i is the degree of firmness with which this information is to be incorporated into the transmuted knowledge system.

`Definition`:
We define a transmutation schema for OEF's, $\star$, to be a function from $\mathcal{E} \times \{\mathbf{L}\backslash\{\bot, \top\}\} \times \mathcal{O}$ to $\mathcal{E}$, where $\mathcal{O}$ is an ordinal, such that $(\mathbf{E}, \mathsf{a}, i) \mapsto \mathbf{E}^\star(\mathsf{a}, i)$ satisfies:

(i) $\mathbf{E}^\star(\mathsf{a}, i)(\mathsf{a}) = i$, and

(ii) $$\mathrm{ks}(\mathbf{E}^\star(\mathsf{a}, i)) = \begin{cases} \{\mathsf{b} \in \mathbf{L} : \ \mathbf{E}(\neg\mathsf{a}) < \mathbf{E}(\neg\mathsf{a} \vee \mathsf{b})\} & \text{if } i > 0 \\ \\ \{\mathsf{b} \in \mathrm{ks}(\mathbf{E}) : \ \mathbf{E}(\mathsf{a}) < \mathbf{E}(\mathsf{a} \vee \mathsf{b})\} & \text{otherwise.} \end{cases}$$

We say $\mathbf{E}^\star(\mathsf{a}, i)$ is an (a, i) – transmutation of $\mathbf{E}$, and we note that a transmutation is not defined for inconsistent, or tautological formulae. Intuitively, a transmutation $\mathbf{E}^\star(\mathsf{a}, i)$, is an OEF where a is assigned the degree of acceptance i, and $\mathrm{ks}(\mathbf{E}^\star(\mathsf{a}, i))$ represents a 'minimal change' of $\mathrm{ks}(\mathbf{E})$.

† The postulates ($^{-}9$) and ($^{*}9$) are essentially required so that the change operator can be modelled by a mapping to the ordinals, that is a well-ordering [9, 10, 14].

In [15] Williams showed, for a finite language $\mathbf{L}$ where $\mathbf{E} \in \mathcal{E}$, i is an ordinal, and $\mathsf{a} \in \mathbf{L}$ is a consistent nontautological formula, that $\mathbf{E}^{\star}(\mathsf{a}, i)$ defined by (Cond – $\mathbf{E}$)‡, below is an (a, i) – transmutation of $\mathbf{E}$.

$$(\text{Cond} - \mathbf{E}) \qquad \mathbf{E}^{\star}(\mathsf{a}, i)(\mathsf{b}) = \begin{cases} -\mathbf{E}(\neg\mathsf{a}) + \mathbf{E}(\mathsf{b}) & \text{if } \mathsf{a} \wedge \neg\mathsf{b} \not\vdash \perp \\ -\mathbf{E}(\mathsf{a}) + \mathbf{E}(\mathsf{b}) + i & \text{otherwise,} \end{cases}$$

where $\mathsf{b} \in \mathbf{L}$ is a dual atom.

We refer to such a transmutation as a conditionalization of $\mathbf{E}$. Spohn [13] has argued that conditionalization is a desirable transmutation, for instance it is reversible and commutative. Furthermore, in [15] Williams showed that for a finite language any transmutation can be achieved via a series of conditionalizations.

Williams [15] also described another transmutation, namely an adjustment. Intuitively, an (a, i)–adjustment is a transmutation which is commanded by the principle of minimal change. That is, $\mathbf{E}$ is changed or disturbed 'as little as necessary' to give $\mathsf{a} \in \mathbf{L}$ the degree of acceptance i.

$$(\text{Adjust–}\mathbf{E}) \qquad \mathbf{E}^{\star}(\mathsf{a}, i) = \begin{cases} \mathbf{E}^{-}(\mathsf{a}) & \text{if } i = 0 \\ (\mathbf{E}^{-}(\mathsf{a}))^{\times}(\mathsf{a}, i) & \text{if } 0 < i < \mathbf{E}(\mathsf{a}) \\ \mathbf{E}^{\times}(\mathsf{a}, i) & \text{otherwise.} \end{cases}$$

where,

$$\mathbf{E}^{-}(\mathsf{a})(\mathsf{b}) = \begin{cases} 0 & \text{if } \mathbf{E}(\mathsf{a}) = \mathbf{E}(\mathsf{a} \vee \mathsf{b}) \\ \mathbf{E}(\mathsf{b}) & \text{otherwise.} \end{cases}$$

$$\mathbf{E}^{\times}(\mathsf{a}, i)(\mathsf{b}) = \begin{cases} 0 & \text{if } \mathbf{E}(\neg\mathsf{a}) = \mathbf{E}(\neg\mathsf{a} \vee \mathsf{b}) \\ \mathbf{E}(\mathsf{b}) & \text{if } \mathbf{E}(\neg\mathsf{a}) < \mathbf{E}(\neg\mathsf{a} \vee \mathsf{b}) \text{ and } \mathbf{E}(\mathsf{b}) > i \\ i & \text{if } \mathbf{E}(\neg\mathsf{a}) < \mathbf{E}(\neg\mathsf{a} \vee \mathsf{b}) \text{ and } \mathbf{E}(\mathsf{b}) \leq i < \mathbf{E}(\neg\mathsf{a} \vee \mathsf{b}) \\ \mathbf{E}(\neg\mathsf{a} \vee \mathsf{b}) & \text{otherwise.} \end{cases}$$

5 Explanation

Spohn [11] gave the following interpretation of the notion of a being a reason for b†.

‡ This condition uses *left-sided subtraction of ordinals*; for ordinals $i \leq j$, $-i + j$ is the uniquely determined ordinal k such that $i + k = j$.

† This section was inspired by the work of Gärdenfors [3] who was the first to attempt to translate Spohn's notion of "reason for" into the AGM paradigm. However, it is not possible to determine the epistemic rank of a formula after a simple contraction and revision operation. Therefore his translation captures the notion of justification rather than "reason for".

($\mathcal{R}1$) $\mathtt{a}$ is a *reason for* $\mathtt{b}$ if and only if raising the epistemic rank of $\mathtt{a}$ would raise the epistemic rank of $\mathtt{b}$.

Accordingly, we can rewrite ($\mathcal{R}1$) using the notion of transmutation, and the degree of acceptance.

($\mathcal{R}2$) $\mathtt{a}$ is a *reason for* $\mathtt{b}$ if and only if increasing the degree of acceptance of $\mathtt{a}$ during a transmutation increases the degree of acceptance of $\mathtt{b}$.

Clearly, $\mathtt{a}$ being a reason for $\mathtt{b}$ is dependent on the transmutation employed.
We can rewrite ($\mathcal{R}2$), using an adjustment as the transmutation, as ($\mathcal{R}3$), below.

($\mathcal{R}3$) $\mathtt{a}$ is a *reason for* for $\mathtt{b}$ if and only if either
(i) $\mathbf{E}(\mathtt{a}) < \mathbf{E}(\mathtt{b})$ and $\mathbf{E}^{\star}(\mathtt{a}, \mathbf{E}(\mathtt{b})+1)(\mathtt{b}) > \mathbf{E}(\mathtt{b})$, or
(ii) $\mathbf{E}(\mathtt{b}) \leq \mathbf{E}(\mathtt{a})$, and $\mathbf{E}^{\star}(\mathtt{a}, \mathbf{E}(\mathtt{a})+1)(\mathtt{b}) > \mathbf{E}(\mathtt{b})$.

Intuitively, ($\mathcal{R}3$) (i) says that if the degree of acceptance of $\mathtt{a}$ is strictly less than the degree of acceptance of $\mathtt{b}$, then if raising the degree of acceptance of $\mathtt{a}$ to $\mathbf{E}(\mathtt{b})+1$ also increases the degree of acceptance of $\mathtt{b}$ then $\mathtt{a}$ is a reason for $\mathtt{b}$. And ($\mathcal{R}3$) (ii), says that if the degree of acceptance of $\mathtt{a}$ is at least as high as the degree of acceptance of $\mathtt{b}$, and increasing the degree of acceptance of $\mathtt{a}$ results in the increase of $\mathtt{b}$'s degree of acceptance, then $\mathtt{a}$ is a reason for $\mathtt{b}$.

`Theorem 1:`
If $\star$ is an adjustment, then ($\mathcal{R}3$) is satisfied if and only if the following hold:

(i) If $\mathbf{E}(\mathtt{a}) < \mathbf{E}(\mathtt{b})$, then $\mathbf{E}(\mathtt{b}) < \mathbf{E}(\neg\mathtt{a} \vee \mathtt{b})$ and $\mathbf{E}(\neg\mathtt{a}) < \mathbf{E}(\neg\mathtt{a} \vee \mathtt{b})$.
(ii) If $\mathbf{E}(\mathtt{b}) \leq \mathbf{E}(\mathtt{a})$, then $\mathbf{E}(\mathtt{a}) < \mathbf{E}(\neg\mathtt{a} \vee \mathtt{b})$ and $\mathbf{E}(\neg\mathtt{a}) < \mathbf{E}(\neg\mathtt{a} \vee \mathtt{b})$.

`Theorem 2:`
If $\star$ is an adjustment, then ($\mathcal{R}3$) is satisfied if and only if $\mathbf{E}(\mathtt{b}) < \mathbf{E}(\neg\mathtt{a} \vee \mathtt{b})$ and $\mathbf{E}(\neg\mathtt{a}) < \mathbf{E}(\neg\mathtt{a} \vee \mathtt{b})$.

In [13] Spohn revised his notion of reason for, and gave a condition which is equivalent [15] to ($\mathcal{R}4$) below, this condition is related to conditionalization.

($\mathcal{R}4$) $\mathtt{a}$ is a *reason for* $\mathtt{b}$ if and only if
(i) $-\mathbf{E}(\neg\mathtt{a}) + \mathbf{E}(\neg\mathtt{a} \vee \mathtt{b}) > -\mathbf{E}(\mathtt{a}) + \mathbf{E}(\mathtt{a} \vee \mathtt{b})$, or
(ii) $-\mathbf{E}(\mathtt{a}) + \mathbf{E}(\mathtt{a} \vee \neg\mathtt{b}) > -\mathbf{E}(\neg\mathtt{a}) + \mathbf{E}(\neg\mathtt{a} \vee \neg\mathtt{b})$.

Intuitively, this condition says that $\mathtt{b}$ has a higher degree of acceptance, or degree of plausibility given $\mathtt{a}$ rather than given $\neg\mathtt{a}$ [13]. However, we shall show that this condition leads to counterintuitive results using `Example 1` in the next section.
Using the notion of "reason for" we define the notion of explanation, and a most plausible explanation, where the most plausible explanation is an explanation which is capable of increasing the degree of acceptance of the explanandum† as much as any other explanation

† The fact to be explained.

during a transmutation.

`Definition:`

We define A to be a set of explanations for $\mathtt{b}$, if every element of A is a reason for $\mathtt{b}$ and not logically equivalent to $\mathtt{b}$.

`Definition:`

Let A be a set of explanations for a formula $\mathtt{b}$. Define $\mathtt{a} \in A$ to be a *most plausible explanation* for $\mathtt{b}$ if $\mathbf{E}^\star(\mathtt{c}, i)(\mathtt{b}) \leq \mathbf{E}^\star(\mathtt{a}, i)(\mathtt{b})$ for all $\mathtt{c} \in A$, and for all ordinals i.

`Theorem 3:`

Let A be a set of explanations for a formula $\mathtt{b}$. Let the transmutation be an adjustment. Then $\mathtt{a}$ is a most plausible explanation in A for $\mathtt{b}$ if and only if

$$\mathbf{E}(\mathtt{c} \to \mathtt{b}) \leq \mathbf{E}(\mathtt{a} \to \mathtt{b}) \text{ for all } \mathtt{c} \in A.$$

6 Examples

As an example we take the discovery of pulsars. The following propositions are relevant:

- `a`: *we are receiving a periodic signal from beyond the solar system*
- `b`: *a spinning neutron star is a remnant of a supernova*
- `c`: *the Crab Nebula marks the site of a supernova*
- `d`: *the source of the signal is a planet*
- `e`: *the source of the signal is inside the solar system*
- `f`: *there is a pulsar within the Crab Nebula*
- `g`: *the signal is a message from an advanced extraterrestial civilization*
- `h`: *the signal is the transmission from man-made space probes*
- `i`: *the signal is composed of pulses emitted from a spinning neutron star*
- `j`: *the signal is composed of pulses emitted from mode-locked masers formed by guided waves propagating around white dwarf stars*
- `k`: $E = mc^2$

The proposition $\mathtt{a}$ represents an observation while the other propositions may constitute explanations for this observation. In the following examples we shall determine which are considered explanations and, in particular, those most plausible explanations according to the framework outlined above.

`Example 1:`

The knowledge set represented by $\mathbf{E}$ holds that the signal is coming from beyond the solar system†, it is not known whether the signal is coming from a planetary object or a star, and the existence of a pulsar in the Crab Nebula has not been discovered yet. More precisely, we have the following:

Let $\mathbf{E} \in \mathcal{E}$ be such that $\mathbf{E}(\mathtt{d}) = 0$, $\mathbf{E}(\mathtt{e}) = 0$, $\mathbf{E}(\mathtt{f}) = 0$, $\mathbf{E}(\mathtt{g}) = 0$, $\mathbf{E}(\mathtt{h}) = 0$, $\mathbf{E}(\mathtt{i}) = 0$, $\mathbf{E}(\mathtt{j}) = 0$ $\mathbf{E}(\neg\mathtt{a}) = 0$, $\mathbf{E}(\neg\mathtt{b}) = 0$, $\mathbf{E}(\neg\mathtt{c}) = 0$, $\mathbf{E}(\neg\mathtt{d}) = 0$, $\mathbf{E}(\neg\mathtt{e}) = 0$, $\mathbf{E}(\neg\mathtt{f}) = 0$, $\mathbf{E}(\neg\mathtt{g}) = 0$, $\mathbf{E}(\neg\mathtt{h}) = 0$, $\mathbf{E}(\neg\mathtt{i}) = 0$, $\mathbf{E}(\neg\mathtt{j}) = 0$

† The absence of any parallax greater than about two arc minutes shows that the source lies far outside the solar system.

$\mathbf{E}(\mathtt{a}) = 5$, $\mathbf{E}(\mathtt{e} \to \mathtt{a}) = 5$, $\mathbf{E}(\mathtt{h} \to \mathtt{a}) = 5$, $\mathbf{E}(\mathtt{k} \to \mathtt{a}) = 5$
$\mathbf{E}(\mathtt{d} \to \mathtt{a}) = 6$, $\mathbf{E}(\mathtt{i} \to \mathtt{a}) = 6$, $\mathbf{E}(\mathtt{j} \to \mathtt{a}) = 6$
$\mathbf{E}(\mathtt{b}) = 7$
$\mathbf{E}(\mathtt{g} \to \mathtt{a}) = 8$
$\mathbf{E}(\mathtt{c}) = 10$
$\mathbf{E}(\mathtt{k}) = 20$, $\mathbf{E}(\mathtt{g} \to \mathtt{d}) = 20$, $\mathbf{E}(\mathtt{h} \to \mathtt{e}) = 20$
If $\star$ is an adjustment of $\mathbf{E}$, then we obtain the following:
$\mathbf{E}^\star(\mathtt{h}, 6)(\mathtt{a}) = 5$, that is, $\mathbf{E}^\star(\mathtt{h}, 6)(\mathtt{a}) = \mathbf{E}(\mathtt{a})$, therefore $\mathtt{h}$ is not an explanation for $\mathtt{a}$. Similarly, $\mathtt{e}$ and $\mathtt{k}$ are not explanations for $\mathtt{a}$.
However, $\mathbf{E}^\star(\mathtt{d}, 6)(\mathtt{a}) = 6$, $\mathbf{E}^\star(\mathtt{g}, 6)(\mathtt{a}) = 6$, $\mathbf{E}^\star(\mathtt{i}, 6)(\mathtt{a}) = 6$, $\mathbf{E}^\star(\mathtt{j}, 6)(\mathtt{a}) = 6$. Therefore, $\mathtt{d}$, $\mathtt{g}$ $\mathtt{i}$ and $\mathtt{j}$ are plausible explanations for $\mathtt{a}$. By Theorem 3, $\mathtt{g}$ is the most plausible explanation for $\mathtt{a}$. We define a *most specific* explanation in the next section, and it turns out that since $\mathtt{g} \to \mathtt{d}$, $\mathtt{g}$ is also a most specific explanation.

If we use $(\mathcal{R}4)$ as our definition of a reason for then we obtain, the following:
$\mathbf{E}(\mathtt{d}) = 0$, $\mathbf{E}(\neg\mathtt{d}) = 0$, $\mathbf{E}(\neg\mathtt{d} \vee \mathtt{a}) = 6$, $\mathbf{E}(\mathtt{d} \vee \mathtt{a}) = 5$, $\mathbf{E}(\mathtt{d} \vee \neg\mathtt{a}) = 0$, $\mathbf{E}(\neg\mathtt{d} \vee \neg\mathtt{a}) = 0$.
Since $\mathbf{E}(\neg\mathtt{d} \vee \mathtt{a}) > \mathbf{E}(\mathtt{d} \vee \mathtt{a})$ we have that $\mathtt{d}$ is an explanation for $\mathtt{a}$.
$\mathbf{E}(\mathtt{e}) = 0$, $\mathbf{E}(\neg\mathtt{e}) = 0$, $\mathbf{E}(\neg\mathtt{e} \vee \mathtt{a}) = 5$, $\mathbf{E}(\mathtt{e} \vee \mathtt{a}) = 5$, $\mathbf{E}(\mathtt{e} \vee \neg\mathtt{a}) = 0$, $\mathbf{E}(\neg\mathtt{e} \vee \neg\mathtt{a}) = 0$,
Therefore $\mathtt{e}$ is not an explanation for $\mathtt{a}$.
$\mathbf{E}(\mathtt{g}) = 0$, $\mathbf{E}(\neg\mathtt{g}) = 0$, $\mathbf{E}(\neg\mathtt{g} \vee \mathtt{a}) = 8$, $\mathbf{E}(\mathtt{g} \vee \mathtt{a}) = 5$, $\mathbf{E}(\mathtt{g} \vee \neg\mathtt{a}) = 0$, $\mathbf{E}(\neg\mathtt{g} \vee \neg\mathtt{a}) = 0$.
Since $\mathbf{E}(\neg\mathtt{g} \vee \mathtt{a}) > \mathbf{E}(\mathtt{g} \vee \mathtt{a})$ we have that $\mathtt{g}$ is an explanation for $\mathtt{a}$.
To illustrate a counter-intuitive explanation we use $\mathtt{k}$, which turns out to be an explanation simply by virtue of the fact that it is strictly more entrenched than $\mathtt{a}$.
$\mathbf{E}(\mathtt{k}) = 20$, $\mathbf{E}(\neg\mathtt{k}) = 0$, $\mathbf{E}(\neg\mathtt{k} \vee \mathtt{a}) = 5$, $\mathbf{E}(\mathtt{k} \vee \mathtt{a}) = 20$, $\mathbf{E}(\mathtt{k} \vee \neg\mathtt{a}) = 20$, $\mathbf{E}(\neg\mathtt{k} \vee \neg\mathtt{a}) = 0$.
Since $-\mathbf{E}(\neg\mathtt{k}) + \mathbf{E}(\neg\mathtt{k} \vee \mathtt{a}) > -\mathbf{E}(\mathtt{k}) + \mathbf{E}(\mathtt{k} \vee \mathtt{a})$, we have that $\mathtt{k}$ is an explanation for $\mathtt{a}$.

`Example 2:`
The knowledge set represented by $\mathbf{E}$, below, holds that the signal is coming from beyond the solar system, the signal is not coming from a planetary object‡, and the existence of a pulsar in the Crab Nebula has not been discovered yet.
Let $\mathbf{E} \in \mathcal{E}$ be such that
$\mathbf{E}(\mathtt{d}) = 0$, $\mathbf{E}(\mathtt{e}) = 0$, $\mathbf{E}(\mathtt{f}) = 0$, $\mathbf{E}(\mathtt{g}) = 0$, $\mathbf{E}(\mathtt{h}) = 0$, $\mathbf{E}(\mathtt{i}) = 0$, $\mathbf{E}(\mathtt{j}) = 0$,
$\mathbf{E}(\neg\mathtt{a}) = 0$, $\mathbf{E}(\neg\mathtt{b}) = 0$, $\mathbf{E}(\neg\mathtt{c}) = 0$, $\mathbf{E}(\neg\mathtt{d}) = 0$, $\mathbf{E}(\neg\mathtt{e}) = 0$, $\mathbf{E}(\neg\mathtt{f}) = 0$, $\mathbf{E}(\neg\mathtt{g}) = 0$, $\mathbf{E}(\neg\mathtt{h}) = 0$, $\mathbf{E}(\neg\mathtt{i}) = 0$, $\mathbf{E}(\neg\mathtt{j}) = 0$
$\mathbf{E}(\mathtt{a}) = 5$, $\mathbf{E}(\mathtt{d} \to \mathtt{a}) = 5$, $\mathbf{E}(\mathtt{e} \to \mathtt{a}) = 5$, $\mathbf{E}(\mathtt{g} \to \mathtt{a}) = 5$, $\mathbf{E}(\mathtt{h} \to \mathtt{a}) = 5$, $\mathbf{E}(\mathtt{k} \to \mathtt{a}) = 5$
$\mathbf{E}(\mathtt{i} \to \mathtt{a}) = 6$, $\mathbf{E}(\mathtt{j} \to \mathtt{a}) = 6$
$\mathbf{E}(\mathtt{b}) = 7$
$\mathbf{E}(\mathtt{c}) = 10$
$\mathbf{E}(\mathtt{k}) = 20$, $\mathbf{E}(\mathtt{g} \to \mathtt{d}) = 20$, $\mathbf{E}(\mathtt{h} \to \mathtt{e}) = 20$
If $\star$ is an adjustment of $\mathbf{E}$, then we obtain the following:

‡ The absence of any Doppler variation in the pulse repetition rate (beyond that ascribable to the earth's orbital motion) makes it unlikely that the source of emission is located on a planet.

$\mathbf{E}^\star(\mathtt{g},6)(\mathtt{a}) = 5$, that is, $\mathbf{E}^\star(\mathtt{g},6)(\mathtt{a}) = \mathbf{E}(\mathtt{a})$, and similarly for $\mathtt{h}$, $\mathtt{k}$, $\mathtt{e}$, $\mathtt{d}$. Therefore, $\mathtt{g}$, $\mathtt{h}$, $\mathtt{k}$, $\mathtt{e}$ and $\mathtt{d}$ are not plausible explanations for $\mathtt{a}$. However $\mathbf{E}^\star(\mathtt{i},6)(\mathtt{a}) = 6$, that is, $\mathbf{E}^\star(\mathtt{i},6)(\mathtt{a}) > \mathbf{E}(\mathtt{a})$, and similarly for $\mathtt{j}$. Hence $\mathtt{i}$ and $\mathtt{j}$ are explanations for $\mathtt{a}$, and by Theorem 3, they are equally plausible explanations for $\mathtt{a}$.

`Example 3:`
The knowledge set represented by $\mathbf{E}$, below, holds that the signal is coming from beyond the solar system, the signal is not coming from a planetary object, and a pulsar in the Crab Nebula has been discovered.

$\mathbf{E}(\mathtt{d}) = 0,\ \mathbf{E}(\mathtt{e}) = 0,\ \mathbf{E}(\mathtt{g}) = 0,\ \mathbf{E}(\mathtt{h}) = 0,\ \mathbf{E}(\mathtt{i}) = 0,\ \mathbf{E}(\mathtt{j}) = 0$

$\mathbf{E}(\neg\mathtt{a}) = 0,\ \mathbf{E}(\neg\mathtt{b}) = 0,\ \mathbf{E}(\neg\mathtt{c}) = 0,\ \mathbf{E}(\neg\mathtt{d}) = 0,\ \mathbf{E}(\neg\mathtt{e}) = 0,\ \mathbf{E}(\neg\mathtt{f}) = 0,\ \mathbf{E}(\neg\mathtt{g}) = 0,\ \mathbf{E}(\neg\mathtt{h}) = 0,\ \mathbf{E}(\neg\mathtt{i}) = 0,\ \mathbf{E}(\neg\mathtt{j}) = 0$

$\mathbf{E}(\mathtt{a}) = 5,\ \mathbf{E}(\mathtt{f}) = 5,\ \mathbf{E}(\mathtt{d} \to \mathtt{a}) = 5,\ \mathbf{E}(\mathtt{e} \to \mathtt{a}) = 5,\ \mathbf{E}(\mathtt{g} \to \mathtt{a}) = 5,\ \mathbf{E}(\mathtt{h} \to \mathtt{a}) = 5,\ \mathbf{E}(\mathtt{k} \to \mathtt{a}) = 5$

$\mathbf{E}(\mathtt{j} \to \mathtt{a}) = 6$

$\mathbf{E}(\mathtt{b}) = 7,\ \mathbf{E}(\mathtt{i} \to \mathtt{a}) = 7$

$\mathbf{E}(\mathtt{c}) = 10$

$\mathbf{E}(\mathtt{k}) = 20,\ \mathbf{E}(\mathtt{g} \to \mathtt{d}) = 20,\ \mathbf{E}(\mathtt{h} \to \mathtt{e}) = 20$

$\mathbf{E}^\star(\mathtt{g},6)(\mathtt{a}) = 5$ hence $\mathbf{E}^\star(\mathtt{g},6)(\mathtt{a}) = \mathbf{E}(\mathtt{a})$, similarly for $\mathtt{h}$, $\mathtt{k}$, $\mathtt{e}$, $\mathtt{d}$. Therefore, $\mathtt{g}$, $\mathtt{h}$, $\mathtt{k}$, $\mathtt{e}$ and $\mathtt{d}$ are not plausible explanations for $\mathtt{a}$, as in the previous example. However $\mathbf{E}^\star(\mathtt{i},6)(\mathtt{a}) = \mathbf{E}^\star(\mathtt{j},6)(\mathtt{a}) = 6 > \mathbf{E}(\mathtt{a})$, and by Theorem 3, $\mathtt{i}$ is a more plausible explanation than $\mathtt{j}$ for $\mathtt{a}$.

7 Abductive Reasoning

One method currently gaining popularity (especially within the field of artificial Intelligence) for defining the notion of explanation is that of abductive reasoning (or simply abduction) [6,7,13]. Abduction is a form of logical inference that aims to derive plausible explanations for data and is, in fact, often described as "inference to the best explanation". The term "abduction" was first coined by the American philosopher Charles Sanders Peirce though its origins can be traced back to Aristotle. An abductive inference proceeds by "proposing" or generating hypotheses which would account or "explain" a group of facts. A common way of defining explanation as an abductive inference is the following (see [6], for example).

`Definition`:
An abductive explanation of a formula $\mathtt{o}$ with respect to a domain theory Γ is a formula $\mathtt{e}$ such that

(i) $\Gamma \cup \{\mathtt{e}\} \vdash \mathtt{o}$, and
(ii) $\Gamma \cup \{\mathtt{e}\} \nvdash \bot$

Within the framework of transmutations described in the preceding pages, our domain theory is taken to be the knowledge set $\mathrm{ks}(\mathbf{E})$. Therefore, we obtain the following definition.

`Definition`:
An abductive explanation of a formula $\mathtt{o}$ with respect to a knowledge set $\mathrm{ks}(\mathbf{E})$ is a formula

$\mathtt{e}$ such that

(i) $\mathrm{ks}(\mathbf{E}) \cup \{\mathtt{e}\} \vdash \mathtt{o}$, and

(ii) $\mathrm{ks}(\mathbf{E}) \cup \{\mathtt{e}\} \nvdash \perp$

Part (i) of the definition corresponds to having the formula $\mathtt{o}$ accepted whenever $\mathtt{e}$ is added to the current knowledge set, while (ii) restricts $\mathtt{e}$ to being consistent with the current knowledge set. In other words, $\mathtt{e}$ is an explanation for $\mathtt{o}$ if $\mathtt{o} \in \mathrm{ks}(\mathbf{E}^\star(\mathtt{e}, i))$ for some ordinal i, provided $\neg\mathtt{e} \notin \mathrm{ks}(\mathbf{E})$.

The notion of "reason for" and subsequent notion of explanation presented here in terms of adjustment (i.e., condition ($\mathcal{R}3$)) will not always correspond to that of abductive explanation. An explanation in terms of adjustment may violate the consistency condition (ii) required of an abductive explanation. For example, if we modify `Example 1` slightly so that $\neg\mathtt{d}$ is accepted initially (say, $\mathbf{E}(\neg\mathtt{d}) = 3$, $\mathbf{E}(\mathtt{d}) = 0$) then we will find that $\mathtt{d}$ is still an explanation for $\mathtt{a}$†. The reason for this is that abduction (as defined above) is basically a process of expansion whilst a transmutation involves the more general process of revision. However, we can show that if $\mathtt{a}$ is a reason for $\mathtt{b}$ according to ($\mathcal{R}3$) and $\mathbf{E}(\neg\mathtt{a}) = 0$ then $\mathtt{a}$ will be an abductive explanation of $\mathtt{b}$ with respect to $\mathrm{ks}(\mathbf{E})$. For a similar reason, explanation in terms of conditionalisation ($\mathcal{R}4$) will not always coincide with abductive explanation.

Conversely, an abductive explanation may not always be an ($\mathcal{R}3$) adjustment explanation. In `Example 1`, for instance, $\mathtt{h}$ was not an explanation for $\mathtt{a}$ (since it did not raise $\mathtt{a}$'s degree of acceptance) yet $\mathtt{h}$ would be an abductive explanation for $\mathtt{a}$ (since $\mathtt{h} \to \mathtt{a} \in \mathrm{ks}(\mathbf{E})$ and $\neg\mathtt{h} \notin \mathrm{ks}(\mathbf{E})$). Also, the definition of abductive explanation above does not rule out the trivial explanation — $\mathtt{a}$ is an explanation for $\mathtt{a}$ — while our definition of explanation in terms of adjustment does not allow this.

The definition of explanation, through the process of abductive inference, will in general classify many formulae as explanations. Often we are interested in determining a single "best" explanation or, at the very least, reducing the number of possible explanations. This is usually done by adopting some form of preference criteria (e.g., cost-based measures [14], probability [7], etc.,). In the framework discussed here we are already equipped with a preference ordering over the formulas — namely, an OEF which assigns a degree of acceptance to each formula. Given this interpretation of OEF's it would seem natural to adopt this ordering to characterise best explanations and this is the essence of section 5.

Another important consideration when determining abductive explanations is the "specificity" of the explanation. The notion of most (and least) specific abduction was introduced by Stickel [14]. It is described within a resolution-based system for performing abduction. A most specific abduction only allows pure literals (literals that resolve with no other clause in the knowledge base) to be used in an abduction. He suggests that most specific abductions are suited to diagnostic tasks while least specific abductions are more appropriate for natural language interpretation. We present a more general definition which

† Note that this does not mean that $\mathbf{E}^\star(\mathtt{d}, i)$ is inconsistent since $\mathbf{E}^\star(\mathtt{d}, i)(\neg\mathtt{d}) = 0$. That is, $\neg\mathtt{d}$ will not be accepted in the resulting knowledge set.

allows us to compare the specificity of certain explanations and then apply this to our ($\mathcal{R}3$) condition.

`Definition`:

An abductive explanation $\mathtt{e}$ of $\mathtt{o}$ with respect to $\mathrm{ks}(\mathbf{E})$ is more specific than an abductive explanation $\mathtt{e}'$ of $\mathtt{o}$ with respect to $\mathrm{ks}(\mathbf{E})$ if and only if $\mathrm{ks}(\mathbf{E}) \cup \{\mathtt{e}\} \vdash \mathtt{e}'$.

Applying this to our notion of explanation in terms of adjustment ($\mathcal{R}3$) we obtain:

`Definition`:

Let A be a set of explanations for a formula $\mathtt{c}$. Define $\mathtt{a} \in A$ to be a more specific explanation for $\mathtt{c}$ than $\mathtt{b} \in A$ if and only if $\mathtt{a}$ is a reason for $\mathtt{b}$. Furthermore, if there is no other formula $\mathtt{b} \in A$ which is a more specific explanation for $\mathtt{c}$ than $\mathtt{a}$, then $\mathtt{a}$ is also called a *most specific* explanation for $\mathtt{c}$.

For instance, in `Example 1` both $\mathtt{g}$ and $\mathtt{d}$ are reasons for $\mathtt{a}$ but $\mathtt{g}$ is a reason for $\mathtt{d}$ and therefore is the more specific (in fact, a most specific) explanation.

`Theorem 2` allows us to say that if $\mathtt{a}$ and $\mathtt{b}$ are both explanations for $\mathtt{c}$ and $\mathtt{a}$ is also an explanation for $\mathtt{b}$ then the more specific explanation $\mathtt{a}$ is more plausible if $\mathbf{E}(\mathtt{a} \to \mathtt{c}) \leq \mathbf{E}(\mathtt{b} \to \mathtt{c})$. This can be generalised for longer "chains" of explanation to yield a method for determining which level of specificity is preferable.

8 Discussion

Transmutations of knowledge systems describe not only how its set of knowledge is revised, but how its underlying preference structure is revised. Hence we are able to capture Spohn's notion of explanation such that it is dependent on the particular transmutation employed. An adjustment is a 'minimal transmutation', that is, it is a transmutation in which the underlying preference relation is changed as little as necessary to effect a change in view. We explored the use of explanation based on adjustment, and moreover, we showed how an adjustment can be used to determine a most plausible explanation, by selecting an explanation that is capable of raising the epistemic rank of the explanandum the most.

We have also contrasted this definition of explanation in terms of adjustment with a common definition of explanation in terms of abductive reasoning. Furthermore, the notion of specificity, used in abductive reasoning, was adapted to our framework and we noted when a more specific explanation is considered a more plausible explanation.

Acknowledgements

The authors would like to thank their colleagues from the Knowledge Systems Group whose comments have improved the content and presentation in this paper. The second author is supported by an Australian Postgraduate Research Award.

References

1. Alchourrón, C., Gärdenfors, P., Makinson, D., *On the Logic of Theory Change: Partial Meet Functions for Contraction and Revision*, Journal of Symbolic Logic, 50:

510-530, 1985.

2. Gärdenfors, P., *Knowledge in Flux: Modeling the Dynamics of Epistemic States*, Bradford Books, MIT Press, Cambridge Massachusetts, 1988.
3. Gärdenfors, P., *The Dynamics of Belief Systems: Foundations vs. Coherence Theories*, Revue Internationale de Philosophie, 44: 24 – 46, 1990.
4. Gärdenfors, P., and Makinson, D., *Revisions of Knowledge Systems using Epistemic Entrenchment*, Proceedings of the Second Conference on Theoretical Aspects of Reasoning about Knowledge, p83 – 96, 1988.
5. Grove, A., *Two modellings for Theory Change*, Journal of Philosophical Logic 17: 157 – 170, 1988.
6. Jackson, P., *Propositional Abductive Logic*, Proceedings of the Society for the Study of Artificial Intelligence and Simulation of Behaviour, p89 – 94, 1989.
7. Peng, Y., and Reggia, J.A., *Abductive Inference Models for Diagnostic Problem-Solving*, Springer-Verlag, 1990.
8. Peppas, P., and Williams, M.A., *A Unified View of the Constructive Modellings for Revision*, Technical Report No. 431, University of Sydney, March 1992.
9. Peppas, P., and Wobcke, W., *On the Use of Epistemic Entrenchment in Reasoning about Action*, European Conference on Artificial Intelligence, p403 – 407, 1992.
10. Peppas, P., *Belief Change and Reasoning about Action: An Axiomatic Approach to Modelling Inert Dynamic Worlds and the Connection to the logic of Theory Change*, PhD Thesis, The University of Sydney, Australia, in preparation.
11. Rott, H., *On the Logic of Theory Change: Partial Meet Contraction and Prioritized Base Contraction*, Report 27, Zentrum Philosophie and Wissenschaftheorie, Universität Konstanz, 1992.
12. Spohn, W., *Deterministic and Probabilistic Reasons and Causes*, Erkenntnis 19: 371-396, 1983.
13. Spohn, W., *Ordinal Conditional Functions: A Dynamic Theory of Epistemic States*, In Harper, W.L., and Skyrms, B. (eds), Causation in decision, belief change, and statistics, II, Kluwer Academic Publishers, p105 – 134, 1988.
14. Stickel, M.E., *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.
15. Williams, M.A., *Transmutations of Knowledge Systems*, PhD Thesis, University of Sydney, 1993, submitted.