A study on the relation between *explanans* and *explanandum*: a modal approach

Marta Cialdea Mayer* & Fiora Pirri†

1 Introduction

Abduction is a natural form of reasoning in any domain where new knowledge may be gained by observing real world facts. It is also a common form of everyday reasoning and its formalization has recently attracted much attention in the AI community. In fact, abduction plays an important role in intelligent problem solving tasks such as diagnosis, natural language interpretation, plan recognition and scene interpretation.

An abductive inference is performed when new evidence informs the knowledge of the reasoner; from the fact that a certain hypothesis would explain the evidence, the truth of this hypothesis is inferred. In general, there may be different hypotheses explaining the same evidence, so alternative ones have to be rejected before the inference is performed. In other terms, the inference is allowed when one is convinced that no other explanation is plausible enough to be accepted.

The characterization of the relation between an observation (the *explanandum*) and a plausible reason that explains it (the *explanans*) is a problem that deserves investigation in its own right, independently of whether or not the inference of the *explanans* is actually performed. In fact, the alternative hypotheses can be compared according to different criteria and different features can be considered such as simplicity, plausibility, capacity of explanation, being more or less *ad hoc*, making inessential assumptions, etc. The debate in the area of Philosophy of Science is old (see, for example [Har78, Pei62, Tha78]). In this tradition the way an hypothesis is judged often depends on the context, i.e. the background theory presently adopted by the reasoning agent: two *explanantes* are compared taking into account the theories that would be obtained if they were adopted. The formal approach to explanation given in this tradition

*Dipartimento di Informatica, Terza Università di Roma, via della Vasca Navale 84, 00146 Roma, Italia. e-mail: cialdea@inf.uniroma3.it

†Dipartimento di Informatica, Università di Roma "La Sapienza", via Salaria 113, 00198 Roma, Italia. e-mail: pirri@dis.uniroma1.it

mainly relies on inductive inference and statistical explanation.

Along the lines of [Hem62] and [G$\ddot{9}$0], Boutilier and Becher [BB93a, BB93b] give a formal logical account of explanation in terms of belief revision. They capture the notion of explanation through a conditional connective $A \Rightarrow B$, and deal with abduction as a revision process of the state of belief of an agent. This approach accounts for the authors' aim of formalizing a notion where it is not required that an *explanans* deductively entail the *explanandum*, nor is it assumed that the observation is consistent with the background knowledge (belief set). Preferred *explanantes* are those that require the least change in the belief set in order to be accepted. Therefore, also in this case, preference depends on the background theory.

However, much research in abductive reasoning relies on a different notion of abduction (see for instance [Lev89, Poo88, RdK87]). In the Artificial Intelligence tradition, in fact, the characterization of explanation has been connected to the generative problem of computing *explanantes* and the research has been focused mainly on the study of proof-theoretical methods for the generation of possible *explanantes*, that are subsequently compared and filtered to obtain the "good" ones (see [CP86, DFndC91]). These works make some assumptions that simplify the generative task: it is assumed that the new evidence B to be explained is consistent with the background theory Θ and that any *explanans* of B deductively entails B in Θ. Howeover, the preference relation between formulae and the criteria for the best choice of the *explanans* are given a logical formulation that is independent of the background theory Θ. In spite of these simplifying assumption, a purely logical characterization of this kind of abductive relation (i.e. the relation between a "good" *explanans* and the *explanandum*) seems harder to obtain and the existing formulations all rely on meta-logical notions.

In particular, because obviously material implication is inadequate to capture the notion of causality that is somewhat implicit in the informal notion of explanation, additional external (meta-logical) conditions are imposed on *explanantes*, with the aim of selecting "good reasons" for an *explanandum* B among all the formulae that logically imply B in the background theory Θ. A key role in characterizing such good reasons is played by the notion of minimality. An abductive *explanans* is required to be minimal in the set of all the formulae that logically entail the *explanandum*, and minimality is defined with respect to a partial order of formulae: a "preference criterium", that allows two potential *explanantes* to be compared in order to determine whether one of them is better than the other. Preference criteria have been stated in many different ways, from absolutely general to very empirical ones. In [Lev89] preference is defined as "being simpler", where a (propositional) formula A is simpler than C whenever the set of literals occurring in A is a subset of those occurring in C. A similar notion is used in [Poo88, RdK87]. A straightforward generalization to logics $\mathcal{L}$ that are not classical propositional logic is obtained by establishing that A is preferred to C whenever $C \models_{\mathcal{L}} A$: weaker *explanantes* are preferred[CMP93, CMP95]. In all these works, the comparison does not take into account the background knowledge.

The general notion of abductive relation that is assumed in the above tradition – and that will be called "classical" for short – can be stated as follows. Let $\mathcal{L}$ be a logic and $<$ be a strict partial order on formulae (independent of any theory). Then an abductive problem is a pair $\langle \Theta, B \rangle$, where Θ is a theory and B an *explanandum* consistent with Θ. A is one of the best *explanantes* of $\langle \Theta, B \rangle$ iff not only does A entail B (in Θ), but also there is no formula $C < A$ (i.e. better than A) that entails B (in Θ). Note that syntactical restrictions on the *explanans* – that are usually imposed in order to rule out trivial explanations – can be captured by a suitable formulation of $<$.

If the classical view is equipped with a form of context-dependent preference, establishing that $A < C$ iff $\Theta \models C \rightarrow A$, then the task of recognizing a minimal *explanans* becomes much easier: in fact, it is sufficient to compare a possible *explanans* A with B itself, because the minimal *explanantes* of B in Θ are exactly those formulae A such that $\Theta \models B \equiv A$. Therefore there is no need to compare A with a (possibly infinite) set of alternative *explanantes*.

The same happens in Becher and Boutilier's approach: non trivial cases of preferred *explanantes* A of an *explanandum* B, that is consistent with the background knowledge Θ, are exactly those A that are equivalent to B in Θ. Consequently, under the above conditions, the abductive inference "explaining" B in Θ amounts to inferring B itself. Obviously, in the belief revision context, this fact does not make abduction trivial. In fact, in general, B is not consistent with Θ and therefore the theory Θ has to be contracted to Θ^- so that $\Theta^- \cup \{B\}$ is consistent.

The above considerations show that, although the standard view of abduction is an apparent simplification of the notion considered by Boutilier and Becher, its logical modeling cannot be obtained by means of a straightforward adaptation of their approach.

A formal semantical characterization of the abductive relation in the classical view would clarify the nature of the notion and would lead to the possibility of clarifying the relation between classical abduction and non monotonic inference, as well as specifying the structural rules that govern classical abductive reasoning. This work proposes a formulation of the classical abductive relation that uses a modal language and a Kripke style semantics. The formulation, although very preliminary, is the first attempt to define a logic for classical abduction.

Because in the sequel we refer exclusively to the classical notion of abduction, we shall preferably use the more common term "explanation", instead of *explanans*.

In the rest of this section we sketch the basic ideas informing this work. Let PC be propositional classical logic over a set $\mathbf{P} = \{p, q, ...\}$ of propositonal letters. The space of all possible interpretations of $\mathbf{P}$ is denoted by $\mathbf{S}$. Propositional formulae are denoted by uppercase latin letters and called PC-formulae;

propositional validity is denoted by $\models_{PC}$.

The notion of abductive explanation is usually defined by a statement that involves quantification over PC-formulae: A is a "good" reason for B in the context of the theory Θ if $\Theta \models_{PC} A \rightarrow B$ and there exists no C such that $\Theta \models_{PC} C \rightarrow B$ and $C < A$. Therefore, A is accepted as a good explanation if there are no strictly "better" ones in the set of those formulae C's such that $\Theta \models_{PC} C \rightarrow B$.

The basic intuition underlying our work comes from the fact that the quantification over formulae and the ordering relation among them suggest a way of modelling abduction by means of Kripke style semantics, i.e. a set K, partially ordered by a binary ("accessibility") relation $\prec$; the preference relation $<$ between explanations is modelled by means of $\prec$ and quantification over formulae is replaced by quantification over accessible elements of K. Formulae to be ordered should be associated to elements of K in such a way that if A is associated to $k \in K$ and $C < A$, then there exists $k' \in K$ such that $k' \prec k$ and C is associated to k'. Moreover, whenever a formula A is associated to $k \in K$, the language should allow one to express that a certain formula B is a logical consequence of A, i.e. there should be a formula α_B depending only on B such that α_B is "true" in k iff $\models_{PC} A \rightarrow B$. The fact that A is a minimal explanation of B can therefore be expressed in a modal language by saying that for all $k \in K$ such that A is associated to k, α_B is true in k and for all $k' \prec k$, α_B is false in k'.

The association of formulae to elements of K can be carried out by means of structures that model the notion of "only knowing". A modal logic for "only knowing" was defined independently and on different grounds by Humberstone [Hum83, Hum87] and Levesque [Lev90]. Humberstone defines *complementary frames* as those Kripke structures endowed with two binary relations, R and $\overline{R}$, such that $\overline{R}$ is the complement of R. The classical modal language is consequently enriched by a new modal operator, corresponding to the relation $\overline{R}$, and a sound and complete axiom system for complementary structures is given. In the following section we restate Humberstone's results, using conventions and notations from [Bou94].

However, our language is restricted so that the modal operators apply only to *objective* formulae, i.e. classical propositional formulae. In fact, since we are interested in the association of objective formulae with elements of K, a language where no nesting of modal operators is allowed is all that is needed. When the results that follow hold also for the unrestricted language, it will be stated and the corresponding proofs can be found in the Appendix.

Beyond the axioms and rules of Humberstone's system, our logic for weak only knowing contains Levesque's axiom, which characterizes those modal structures containing all classical interpretations. While Levesque's system is based on weak S5, our system si based on the weakest modal system K. Obviously, this "weak" form of only knowing still allows one to express context independent validityno a propositional implication; in fact, in that case, no nesting of modal operators is needed.

The main results of this work apply to languages based on a finite set of propositional letters. In fact, if the language is infinite, strong completeness does not hold. A proof of this fact, together with a proof of weak completeness for an infinite language, is given in the Appendix. However, weak completeness is not enough, in general, to ensure that complete axiomatizations of the order relation among formulae can be given.

2 Weak Only Knowing

In this section we define the system **CKH***. The system is introduced by steps, corresponding to subsystems of **CKH***. They are based on the language that extends a propositional language over a finite set $\mathbf{P}=\{p_1, ..., p_n\}$ of propositional letters by means of the modal operators $\Box$ and $\overleftarrow{\Box}$. The set of *k-formulae* is defined by adding the following clause to the usual inductive definition of the classical propositional language:

- if A is a PC-formula, then $\Box A$ and $\overleftarrow{\Box} A$ are k-formulae.

The dual operators $\Diamond$ and $\overleftarrow{\Diamond}$ are defined as usual in terms of $\Box$ and $\overleftarrow{\Box}$, respectively. We use greek letters to denote modal formulae and reserve latin letters for PC-formulae.

2.1 k-structures and the system KH

A *k-structure* k is a Kripke structure with a distinguished ("actual") world, i.e. $k = \langle w_0, W, R, V\rangle$, where W is a set (of possible worlds), $w_0 \in W$, $R \subseteq W^2$ and $V : W \rightarrow \mathbf{S}$ maps worlds into propositional interpretations of the language $\mathbf{P}$. We shall write wRw' for $\langle w, w'\rangle \in R$ and $w\overline{R}w'$ for $\langle w, w'\rangle \notin R$. As there is no risk of confusion, we shall write $w \in k$ instead of $w \in W$.

A propositional letter p is true in a world $w \in k$ ($k \models_w A$) just in case $V(w)(p) = true$ and the truth of a k-formula α in w, is defined by adding the following clauses to the usual definition:

- $k \models_w \Box A$ iff for all $w' \in k$ such that wRw', $k \models_{w'} A$
- $k \models_w \overleftarrow{\Box} A$ iff for all $w' \in k$ such that $w\overline{R}w'$, $k \models_{w'} A$.

Truth in a k-structure is identified with truth in the actual world w_0, as in Kripke's original semantics:

$$k \models \alpha \text{ iff } w_0 \models \alpha$$

Validity in k-structures is captured by the system **KH** obtained by adding the following axioms and inference rules to PC [Hum83]:[1]

(K) $\Box(\alpha \rightarrow \beta) \rightarrow (\Box\alpha \rightarrow \Box\beta)$ $\quad$ (K') $\overleftarrow{\Box}(\alpha \rightarrow \beta) \rightarrow (\overleftarrow{\Box}\alpha \rightarrow \overleftarrow{\Box}\beta)$

(Nec) $\dfrac{\alpha}{\Box\alpha}$ $\quad$ (Nec') $\dfrac{\alpha}{\overleftarrow{\Box}\alpha}$

(H) $(\Box\alpha \wedge \overleftarrow{\Box}\beta) \rightarrow (\alpha \vee \beta)$

In the system **KH** the notion of truth of a PC-formula in every world of a k-structure can be defined by:

$$\overleftrightarrow{\Box}A \equiv_{def} \Box A \wedge \overleftarrow{\Box}A$$

In fact, $R \cup \overline{R} = W^2$, so that $\overleftrightarrow{\Box}$ corresponds to the universal modality (see also [BB93a, GP92]). Anagously:

$$\overleftrightarrow{\Diamond}A \equiv_{def} \Diamond A \wedge \overleftarrow{\Diamond}A$$

2.2 k*-structures and the system KH*

We restrict our attention to k*-structures:

Definition 1 *A k-structure $k = \langle w_0, W, R, V\rangle$ is a* k*-structure *iff* $\{V(w) : w \in W\} = \mathbf{S}$.

The notation $\Gamma \models^* \alpha$ will be used to say that α is true in any k*-structure satisfying every formula in Γ.

Levesque'axiom characterizing k*-structures is

$$\overleftrightarrow{\Diamond}A \qquad \text{for any satisfiable PC-formula } A$$

When the language **P** is finite, the above schema may be replaced by a finite set of axioms, as follows. Let us convene that

Definition 2 *A* name *is a conjunction C of literals such that for every propositional letter $p \in \mathbf{P}$ exactly one between p and $\neg p$ is in C*

[1] If the language is not restricted to formulae whose modal depth is not greater than 1, then in the schema H is to be replaced by:

$$\mathcal{D}(\Box\alpha \wedge \overleftarrow{\Box}\beta) \rightarrow \mathcal{B}(\alpha \vee \beta)$$

where $\mathcal{D}$ is any (possibly empty) sequence of $\Diamond$ and $\overleftarrow{\Diamond}$, and $\mathcal{B}$ is any (possibly empty) sequence of $\Box$ or $\overleftarrow{\Box}$. Therefore, H stands for an infinite set of schemata. An alternative axiomatization can be found in **cit. dal referee**.

The axioms of **KH*** include then the following:

(LP) $\qquad \overleftrightarrow{\Diamond} C \qquad$ for any name C

Obviously, if B is any satisfiable PC-formula, $\overleftrightarrow{\Diamond} B$ is derivable in **KH***. In fact, if $ass : \mathbf{P} \rightarrow \{true, false\}$ is an assigment satisfying B and C is the conjunction of all the literals $p'_i = \begin{cases} p_i & \text{if } ass(p_i) = true \\ \neg p_i & \text{otherwise} \end{cases}$ for $p_i \in \mathbf{P}$, then clearly $\models_{PC} C \rightarrow B$. Since $\overleftrightarrow{\Diamond} C$ is an axiom, $\overleftrightarrow{\Diamond} B$ is derivable.

The system **KH*** is sound and complete with respect to k*-structures. Before establishing this result we introduce some preliminary definitions.

Definition 3

- Γ *is a* PC-maximal consistent set *of PC-formulae iff* Γ *is a consistent set of PC-formulae such that for every PC-formula* $A \notin \Gamma$, $\Gamma \cup \{A\}$ *is propositionally inconsistent.*
- Δ *is a maximal consistent set of propositional literals iff for every* $p \in \mathbf{P}$ *exactly one of* p *and* $\neg p$ *is in* Δ.
- *A set of k-formulae* Γ *is* **KH***-*consistent if* $\Gamma \not\vdash_{KH^*} \bot$.
- Γ *is a* **KH***-maximal consistent set *of formulae iff* Γ *is a* **KH***-*consistent set of k-formulae such that for every k-formula* $\alpha \notin \Gamma$, $\Gamma \cup \{\alpha\}$ *is* **KH***-*inconsistent.*

In the following, if Γ is a finite set of formulae, we shall freely use Γ to denote both the set and the conjunction $\bigwedge_{\gamma \in \Gamma} \gamma$ of the formulae in Γ, since the difference will be clear from the context. Consequently, if ∇ is a modal operator, then $\nabla\Gamma$ denotes $\nabla(\bigwedge_{\gamma \in \Gamma} \gamma)$.

The following lemma is an adaptation to **KH*** of a part of the completeness proof of CO* in [Bou94]. Its proof does not depend on our restriction on the scope of the modal operators, so it holds also when modal operators are allowed to apply to k-formulae, provided Levesque's axiom is usable in its full form. For this reason, both in the statement and proof of the lemma, we use lowercase greek letters to deonote formulae where, in our case, uppercase latin letters denoting objective formulae could be used. Moreover, the lemma holds also for infinite **P**.

Lemma 1 *If* Δ *is a maximal consistent set of propositional literals and* Γ *is a* **KH***-*maximal consistent set of k-formulae, then either* $\Delta \cup \{\alpha \mid \Box\alpha \in \Gamma\}$ *is* **KH***-*consistent, or* $\Delta \cup \{\alpha \mid \overleftarrow{\Box}\alpha \in \Gamma\}$ *is* **KH***-*consistent.*

Proof. Let us assume that $\Delta \cup \{\alpha \mid \Box\alpha \in \Gamma\}$ is not **KH***-consistent. Therefore, some finite subset of $\Delta \cup \{\alpha \mid \Box\alpha \in \Gamma\}$ is not **KH***-consistent and, since $\Box(\alpha_0 \wedge ... \wedge \alpha_n) \in \Gamma$ whenever $\Box\alpha_0, ..., \Box\alpha_n \in \Gamma$, there is some formula $\Box\alpha \in \Gamma$ such that $\Delta_0 \cup \{\alpha\}$ is not **KH***-consistent, where $\Delta_0 \subseteq \Delta$ is finite. Therefore, $\vdash_{KH^*} \alpha \rightarrow \neg\Delta_0$. By (Nec), (K) and PC, $\vdash_{KH^*} \Box\alpha \rightarrow \neg\Box\Delta_0$. Since Γ is maximally consistent, (i) $\Box\neg\Delta_0 \in \Gamma$.

A similar argument shows that if $\Delta \cup \{\alpha \mid \overleftarrow{\Box}\alpha \in \Gamma\}$ is not **KH***-consistent, then there is some formula $\overleftarrow{\Box}\beta \in \Gamma$ and a finite subset $\Delta_1 \subseteq \Delta$, such that $\vdash_{KH^*} \Box\beta \rightarrow \neg\Box\Delta_1$, therefore (ii) $\overleftarrow{\Box}\neg\Delta_1 \in \Gamma$.

Since Γ is maximally consistent, from (i) and (ii) it follows that $\Box\neg(\Delta_0 \wedge \Delta_1) \wedge \overleftarrow{\Box}\neg(\Delta_0 \wedge \Delta_1) \in \Gamma$, i.e. (iii) $\overleftrightarrow{\Box}\neg(\Delta_0 \wedge \Delta_1) \in \Gamma$. However, $\Delta_0 \wedge \Delta_1$ is consistent, therefore, by use of LP, $\overleftrightarrow{\Diamond}(\Delta_0 \wedge \Delta_1) \in \Gamma$, thus contradicting, with (iii), the consistency of Γ. ∎

The system **KH*** is sound and complete with respect to k*-structures, also when the language is infinite and the scope of modal operators is not restricted. The Appendix contains a proof of the general statement, while here we focus on the restricted language that allows a simpler proof. Note that the soundness proof does not depend on our restrictions on the language.

Theorem 1 *The system* **KH*** *is sound and complete with respect to k*-structures, i.e. for any set of k-formulae* $\Gamma \cup \{\alpha\}$, $\Gamma \models^* \alpha$ *iff* $\Gamma \vdash_{KH^*} \alpha$.

Proof. The soundness of the subsystem of **KH*** consisting of K, K', Nec and Nec' is straightforward. The following proof of the soundness of H is exactly like that in [Hum83]: suppose that an instance $\mathcal{D}(\Box\alpha \wedge \overleftarrow{\Box}\beta) \rightarrow \mathcal{B}(\alpha \vee \beta)$ of H is false in some k*-structure $k = \langle w_0, W, R, V\rangle$. Then

$$k \models_{w_0} \mathcal{D}(\Box\alpha \wedge \overleftarrow{\Box}\beta) \quad \text{but} \quad k \not\models_{w_0} \mathcal{B}(\alpha \vee \beta).$$

From the first of these facts, it follows that there is $w_1 \in k$ such that $k \models_{w_1} \Box\alpha \wedge \overleftarrow{\Box}\beta$, i.e. (i) $k \models_{w_1} \Box\alpha$ and (ii) $k \models_{w_1} \overleftarrow{\Box}\beta$; the second implies that there is $w_2 \in k$ such that $k \not\models_{w_2} \alpha \vee \beta$, i.e. (iii) $k \not\models_{w_2} \alpha$ and (iv) $k \not\models_{w_2} \beta$. Since k is complementary, either $w_1 R w_2$ or $w_1 \overline{R} w_2$. But the first case contradicts (i) and (iii), while the second contradicts (ii) and (iv). So, no istance of H can be false in any k*-structure.

Completeness is established as follows, by combining and simplifying the corresponding proofs in [Hum83] and [Bou94]. Let us assume that $\Gamma \not\vdash_{KH^*} \alpha$ and $\Delta \supseteq \Gamma \cup \{\neg\alpha\}$ is a maximally consistent set of formulae. We show how to build a k*-model $k = \langle w_0, W, R, V\rangle$ of Δ. The constrution is carried on as follows.

If S is the set of all the maximal consistent sets of literals in the language $\mathbf{P}$, then we establish $W = \{\Delta\} \cup S$ and $w_0 = \Delta$. Note that, if the language is finite, then W is finite. By Lemma 1, for any element $w \in S$, either $w \cup \{A \mid \Box A \in \Delta\}$ is **KH***-consistent, or $w \cup \{A \mid \overleftarrow{\Box}A \in \Delta\}$ is consistent. In the first case,

$\{A \mid \Box A \in \Delta\}$ is added to w and it is established that $\langle w_0, w\rangle \in R_0$; in the second case $\{A \mid \overleftarrow{\Box} A \in \Delta\}$ is added to w and it is established that $\langle w_0, w\rangle \in \overline{R_0}$. Every $w \in S$ is finally extended to be a PC-maximal consistent set. This completes the first step of the construction.

Then, the following operations are performed:

(a) For every formula $\Diamond A \in \Delta$, add a PC-maximal consistent set w to W, such that $w \supseteq \{A\} \cup \{\beta \mid \Box B \in \Delta\}$; add $\langle w_0, w'\rangle$ to R.

(b) for every formula $\overleftarrow{\Diamond} A \in \Delta$, add a PC-maximal consistent set w to W, such that $w \supseteq \{A\} \cup \{B \mid \overleftarrow{\Box} B \in \Delta\}$; add $\langle w_0, w'\rangle$ to $\overline{R}$.

Note that all the elements in W, but Δ are sets of PC-formulae.

In k, V is such that $V(w)(p) = \mathit{true}$ iff $p \in w$, for all $w \in W$. The fact that for every PC-formula A:

(i) for all $w \in W$, $k \models_w A$ iff $A \in w$

can be proved as usual by induction on A. Moreover, for every k-formula α

(ii) $k \models_{w_0} \alpha$ iff $\alpha \in \Delta$

is proved as follows:

1. If α is a PC-formula, the thesis follows immediately from (i).

2. Let $\alpha = \Box A$. If $\alpha \in \Delta$, then by the construction of k, for all $w \in k$, if $w_0 R w$ then $A \in w$. By (i), for all $w \in k$, if $w_0 R w$ then $k \models_w A$, so $k \models_{w_0} \alpha$. If $\alpha \notin \Delta$, by the maximality of Δ, $\neg\alpha \in \Delta$. By the construction of k, there is a $w \in k$ such that $w_0 R w$ and $\neg A \in w$. Since w is consistent, $A \notin w$, therefore, by the induction hypothesis, $k \not\models_w A$, therefore $k \not\models_{w_0} \alpha$.

3. If $\alpha = \overleftarrow{\Box} A$, the proof is similar to the previous case.

Now we show that k can be extended to a (complementary) k*-structure. It must be shown that it is possible to extend R' and $\overline{R'}$ to R and $\overline{R}$, respectively, so that $R \cup \overline{R} = W^2$. Let $w, w' \in W$ be such that neither $wR'w'$ nor $w\overline{R'}w'$. Assume that it is not possible to add the pair $\langle w, w'\rangle$ to R', consistently with (i), nor is it possible to add the pair $\langle w, w'\rangle$ to $\overline{R'}$. Then there is some formula $\Box A \in w$ such that $A \notin w'$ and there is some $\overleftarrow{\Box} B \in w$ such that $B \notin w'$. If w contains modal formulae, then $w = \Delta$. By construction, for every $w \in W - \{\Delta\}$ either $w_0 R w$ or $w_0 \overline{R} w$. Therefore it must be $w' = \Delta$ too. So, $k \models_{w_0} \Box A \wedge \overleftarrow{\Box} B$ but $k \not\models_{w_0} A \vee B$, thus contradicting (H). ∎

Let k be a k-structure. The notion of truth at *precisely* the accessible points is definable [Hum87, Lev90]:

$$\mathbf{O}\alpha \equiv_{Def} \Box\alpha \wedge \overleftarrow{\Box}\neg\alpha$$

If $k = \langle w_0, W, R, V\rangle$ is a k-structure and $k \models \mathbf{O}\alpha$, then the relation R partitions the elements of the structure into two sets: the set of worlds that are accessible from w_0, where α is true, and the others, where α is false. This fact leads to the following lemma, that establishes that $\mathbf{O}A$ is always satisfiable.

Lemma 2 *If A is a PC-formula, then $\mathbf{O}A$ is satisfiable (in a k*-structure).*

Proof. Let $k = \langle w_0, W, R, V\rangle$ be the k*-structure defined as follows. W is a set having the same cardinality as the set $\mathbf{S}$ of all the possible interpretation of $\mathbf{P}$, w_0 is any element of W. V is any one-to-one mapping from W onto $\mathbf{S}$. $R = \{\langle w_0, w\rangle \mid V(w)(A) = true\}$. We show that $k \models_{w_0} \mathbf{O}A$. Let $w \in W$. If $w_0 R w$, then, by definition, $V(w)(A) = true$, i.e. $k \models_w A$. Therefore $k \models_{w_0} \Box A$. If $w_0 \overline{R} w$, then, by definition, $V(w)(A) = false$, i.e. $k \not\models_w A$, so $k \models_w \neg A$. Therefore $k \models_{w_0} \overleftarrow{\Box}\neg A$. ∎

Definition 4 *If A is a propositional formula, then the k*-model of $\mathbf{O}A$ built in the previous proof is called* the canonical k*-model of $\mathbf{O}A$.

Obviously, if k is any k*-structure, then

$$k \models \overleftrightarrow{\Box}A \text{ iff } \models_{PC} A$$

Therefore:

Lemma 3 *If k is a k*-structure, A a PC-formula and $k \models \mathbf{O}A$, then for every PC-formula B, $k \models \Box B$ iff $\models_{PC} A \rightarrow B$.*

Consequently, by Lemma 2, $\models \mathbf{O}A \rightarrow \Box B$ iff $\models_{PC} A \rightarrow B$. The same property is proved for Levesque's system in [Lev90].

Lemma 4 *Let k be a k*-structure and A and B PC-formulae. If $k \models \mathbf{O}A$ and $k \models \mathbf{O}B$, then $\models_{PC} A \equiv B$.*

The two lemmas above are actually consequences of the following more general statements about k-structures.

Lemma 5 *Let k be a k-structure. If $k \models \mathbf{O}A$, then for every PC-formula B, $k \models \Box B$ iff $k \models \overleftrightarrow{\Box}(A \rightarrow B)$.*

Proof. Let k be a k-structure such that $k \models \mathbf{O}A$, i.e. $k \models \Box A$ and $k \models \overleftarrow{\Box}\neg A$, that is, for all $w \in W$:
- if $w_0 R w$, then $w \models A$, and
- if $w_0 \overline{R} w$, then $w \models \neg A$.

Let us further assume that $k \models \Box B$ for any PC-formula B. Then, for all $w \in W$:
- if $w_0 R w$, then $w \models B$, so that $w \models A \rightarrow B$, and

• if $w_0\overline{R}w$, then $w \models \neg A$, so that $w \models A \rightarrow B$.
Consequently, $k \models \overleftrightarrow{\Box}(A \rightarrow B)$.

Conversely, let us assume $k \models \overleftrightarrow{\Box}(A \rightarrow B)$. Then in particular $k \models \Box(A \rightarrow B)$. By the validity of axiom (K) and the assumption $k \models \Box A$, it follows $k \models \Box B$. ∎

Lemma 6 *Let k be a k-structure. If $k \models \mathbf{O}A$ and $k \models \mathbf{O}B$, then $k \models \overleftrightarrow{\Box}(A \equiv B)$.*

Proof. By lemma 5:
• if $k \models \mathbf{O}A$ and $k \models \Box B$, then $k \models \overleftrightarrow{\Box}(A \rightarrow B)$.
• if $k \models \mathbf{O}B$, and $k \models \Box A$, then $k \models \overleftrightarrow{\Box}(B \rightarrow A)$.
Therefore, for any $w \in W$, $w \models A \equiv B$, i.e. $k \models \overleftrightarrow{\Box}(A \equiv B)$.
∎

The two lemmas above hold in the general language, where the modal operators can be applied to eny formula.

2.3 Canonical k*-structures and the system CKH*

Lemma 4 guarantees the unicity of "only known" formulae, but not their existence. In fact, there are k*-interpetations k such that for no PC-formula A is it the case that $k \models \mathbf{O}A$; this happens, for example, whenever there are $w, w' \in k$ such that $V(w) = V(w')$, w_0Rw and $w_0\overline{R}w'$. This is a feature that distinguishes Humberstone's and Boutilier's semantics from Levesque's semantics. The following definition labels as "canonical" those K*-structures satisfying $\mathbf{O}A$ for some A.

Definition 5 *A k*-structure k is a* canonical k*-structure *iff there is a PC-formula A such that $k \models \mathbf{O}A$.*

If the language $\mathbf{P}$ is infinite, then **KH*** is weakly complete with respect to canonical k*-structures, i.e. for every k-formula α, $\vdash_{KH*} \alpha$ iff α is true in every canonical k*-structure (a proof of this fact is in the Appendix). However, weak completeness is not sufficient for the aims of this work. For this reason the system **CKH*** is based on a finite set $\mathbf{P} = \{p_1, ..., p_n\}$ of propositional letters, since in this case non canonical k*-structures can be ruled out by means of suitable axioms. In fact the formula $\Diamond(p_1 \wedge ... \wedge p_n) \wedge \overleftarrow{\Diamond}(p_1 \wedge ... \wedge p_n)$ is satisfiable in non-canonical k*-structures, but not in canonical ones. Therefore, canonical structures are charaterized by the **KH***-theory consisting of the following set of axioms:

$$\Box(p'_1 \vee ... \vee p'_n) \vee \overleftarrow{\Box}(p'_1 \vee ... \vee p'_n) \qquad \text{where } p'_i \text{ is either } p_i \text{ or } \neg p_i$$

So, the theory **CKH*** of canonical k*-structures over the finite language **P** extends **KH*** by means of:

(FIN) $\Box\neg C \vee \overleftarrow{\Box}\neg C$ for every name C

In [GPT86] a solution similar to ours is adopted, except that names for worlds are new constants instead of maximal consistent conjunctions of literals. The introduction of such new constants, however, is not suitable for our purposes.

Theorem 2 *The system* **CKH*** *is sound and complete w.r.t. canonical k*-structures, i.e. for any set of k-formulae* $\Gamma \cup \{\alpha\}$, $\Gamma \vdash_{CKH^*} \alpha$ *iff for every canonical k*-model* k *of* Γ, $k \models \alpha$.

Proof. We first prove that:

(1) for every k*-structure $k = \langle w_0, W, R, V\rangle$, k is canonical iff for no $w, w' \in W$ it is the case that $V(w) = V(w')$, $w_0 R w$ and $w_0 \overline{R} w'$.

Let $W^+ = \{w \in W \mid w_0 R w\}$ and $W^- = \{w \in W \mid w_0 \overline{R} w\}$ and let us assume that k is canonical, i.e. for some A, $k \models \mathbf{O}A$. If $w \in W^+$ and $w' \in W^-$, then $k \models_w A$ and $k \not\models_{w'} A$; therefore $V(w) \neq V(w')$.

Conversely, assume that for no $w, w' \in W$ it is the case that $V(w) = V(w')$, $w_0 R w$ and $w_0 \overline{R} w'$. Let $\Gamma = \{C \mid C \text{ is a name and for some } w \in W^+,\ k \models_w C\}$ and $A = \bigvee_{C \in \Gamma} C$. Since **P** is finite, Γ is finite, so that A is well defined. Clearly, $k \models \Box A$. If there were $w' \in W^-$ such that $k \models_{w'} A$, then $k \models_{w'} C$ for some $C \in \Gamma$, i.e. there would exist $w \in W^+$ such that $V(w) = V(w')$. But this is excluded by hypothesis, so that $k \models \neg\overleftarrow{\Box} A$. Hence $k \models \mathbf{O}A$.

Now we prove that:

(2) for every k*-structure $k = \langle w_0, W, R, V\rangle$, there exists a name C such that $k \models \Diamond C \wedge \overleftarrow{\Diamond} C$ iff for some $w, w' \in W$ it is the case that $V(w) = V(w')$, $w_0 R w$ and $w_0 \overline{R} w'$.

In fact, if there exists a name C such that $k \models \Diamond C \wedge \overleftarrow{\Diamond} C$, then there are $w \in W^+$ and $w' \in W^-$ such that $k \models_w C$ and $k \models_{w'} C$, i.e. $V(w) = V(w')$. Conversely, let us assume that there are $w \in W^+$ and $w' \in W^-$ such that $V(w) = V(w')$. If $\mathbf{P} = \{p_1, ..., p_n\}$, let us define, for $w \in W$, $C_w = p_1^w \wedge \ldots p_n^w$, where

$$p_i^w = \begin{cases} p_i & \text{if } V(w)(p_i) = true \\ \neg p_i & \text{otherwise} \end{cases}$$

Then $C_w = C_{w'}$ and $k \models \Diamond C_w \wedge \overleftarrow{\Diamond} C_w$.

From (1) and (2) it follows that a k*-structure k is canonical iff it is a model of every instance of *FIN*. The desired result then follows from Theorem 1. ■

In the following section we show how to model a preference relation $<$ on formulae by means a partial order $\prec$ on a multiset K of canonical k*-structures.

3 Preference structures

A preference structure is essentially a modal **K4** structure, where the "points" are canonical k*-structures:

Definition 6 *A* preference structure $\mathcal{M}$ *(P-structure for short) is a structure* $\langle k_0, K, \succ, \varphi \rangle$*, where* K *is a set,* $k_0 \in K$*,* $\succ$ *is a transitive relation on* K *and* φ *is a function that maps every element of* K *to a canonical k*-structure.*

We shall sometimes write $k \in \mathcal{M}$ instead of $k \in K$. Moreover, we shall identify the elements k with the corresponding $\varphi(k)$, treating in fact K as a multiset of k*-structures. Clearly, every $k \in K$ is equipped with its own accessibility relation R_k, so that a p-structure is in fact a multimodal structure. However, as R_k does not interact with the relations in the other k*-structures, this distinction can be ignored.

The language that is interpreted in a P-structure is the language of k-formulae, augmented with the modal operator $[\star]$. P-formulae (or simply formulae) are inductively defined as follows:

(i) if α is a k-formula, then α is a P-formula.

(ii) if α and β are P-formulae, then $\neg\alpha$, $\alpha \vee \beta$, $\alpha \wedge \beta$, $\alpha \rightarrow \beta$, and $[\star]\alpha$ are P-formulae.

Note that a $[\star]$ operator is not allowed to occur inside the scope of a $\Box$ or $\overleftarrow{\Box}$. In fact, k-formulae are the "atoms" over which P-formulae are built in the standard modal way. $\langle\star\rangle\alpha$ is defined as $\neg[\star]\neg\alpha$.

The truth of a formula in $k \in \mathcal{M}$ is defined as follows:

1. if α is a k-formula then $\mathcal{M} \models_k \alpha$ iff $k \models \alpha$
2. $\mathcal{M} \models_k \neg\alpha$ iff $\mathcal{M} \not\models_k \alpha$
3. $\mathcal{M} \models_k \alpha \vee \beta$ iff either $\mathcal{M} \models_k \alpha$ or $\mathcal{M} \models_k \beta$
4. $\mathcal{M} \models_k \alpha \wedge \beta$ iff $\mathcal{M} \models_k \alpha$ and $\mathcal{M} \models_k \beta$
5. $\mathcal{M} \models_k \alpha \rightarrow \beta$ iff either $\mathcal{M} \not\models_k \alpha$ or $\mathcal{M} \models_k \beta$
6. $\mathcal{M} \models_k [\star]\alpha$ iff for all k', if $k \succ k'$, $\mathcal{M} \models_{k'} \alpha$

Truth in $\mathcal{M}$ is defined by:

$$\mathcal{M} \models \alpha \text{ iff } \mathcal{M} \models_{k_0} \alpha$$

From now on, since there is no risk of confusion, the notation $\models$ is relative to preference structures.

The system **AK4** is obtained by adding the following axioms and inference rules to the system **CKH***:

$$(K_P) \qquad [\star](\alpha \rightarrow \beta) \rightarrow ([\star]\alpha \rightarrow [\star]\beta)$$

$$(4) \qquad [\star]\alpha \rightarrow [\star][\star]\alpha$$

$$(Nec_P) \qquad \frac{\alpha}{[\star]\alpha}$$

As the modal operator $[\star]$ does not interact with $\Box$ and $\overleftarrow{\Box}$, **AK4** contains no axioms ruling the connection between the two levels.

AK4 is sound and complete with respect to preference structures. The proof is is a suitable composition of the completeness proofs for **CKH*** and the modal system **K4**.

Lemma 7 *For every set of P-formulae* $\Gamma \cup \{\alpha\}$*, if* $\Gamma \vdash_{AK4} \alpha$*, then* $\Gamma \models \alpha$*.*

Proof. This result easily follows from the soundness of the systems **K4** and **CKH***. ■

Lemma 8 *For every set of P-formulae* $\Gamma \cup \{\psi\}$*, if* $\Gamma \models \psi$*, then* $\Gamma \vdash_{AK4} \psi$*.*

Proof. We first introduce the following definitions. If Γ is a set of P-formulae, Γ is a *propositional expansion* iff:

- if $\alpha \vee \beta \in \Gamma$ then either $\alpha \in \Gamma$ or $\beta \in \Gamma$;
- if $\alpha \wedge \beta \in \Gamma$ then $\alpha \in \Gamma$ and $\beta \in \Gamma$;
- if $\alpha \rightarrow \beta \in \Gamma$ then either $\neg\alpha \in \Gamma$ or $\beta \in \Gamma$;
- if $\neg\neg\alpha \in \Gamma$ then $\alpha \in \Gamma$.
- if $\neg(\alpha \vee \beta) \in \Gamma$ then $\neg\alpha \in \Gamma$ and $\neg\beta \in \Gamma$.
- if $\neg(\alpha \wedge \beta) \in \Gamma$ then either $\neg\alpha \in \Gamma$ or $\neg\beta \in \Gamma$.
- if $\neg(\alpha \rightarrow \beta) \in \Gamma$ then $\alpha \in \Gamma$ and $\neg\beta \in \Gamma$.

Γ' is a propositional expansion of Γ iff Γ' is a propositional expansion, $\Gamma \subseteq \Gamma'$ and no proper subset of Γ' is a propositional expansion. Obviously, if Γ is consistent, there exists a propositional expansion Γ' of Γ that is consistent as well.

If $\Gamma \not\vdash_{AK4} \psi$, we show how to build a P-model $\mathcal{M}$ of $\Gamma \cup \{\neg\psi\}$. We incrementally build a structure $\langle k_0, K, \succ\rangle$, where K is a set of propositional expansions.

At stage 0, K contains the single element k_0 that is any consistent propositional expansion of $\Gamma \cup \{\neg\psi\}$ and $\succ$ is the empty relation.

At stage $i+1$ the following operations are performed. For each new element k added to K at stage i and for each formula $\neg[\star]\alpha \in k$, a new element k' is added to K, that is any consistent propositional expansion of the set $\{\neg\alpha\} \cup \{\beta, [\star]\beta \mid [\star]\beta \in k\}$. That such a consistent expansion exists is due to the fact that if k is consistent, then so is also $\{\neg\alpha\} \cup \{\beta, [\star]\beta \mid [\star]\beta \in k\}$. The element $\langle k, k' \rangle$ is added to $\succ$, as well as all the pairs $\langle k'', k' \rangle$ such that $k'' \succ k$ (i.e. $\succ$ is built up so as to be transitive).

The elements of K are all consistent sets of P-formulae. For every $k \in K$, let $\xi(k)$ denote the subset of k containing only k-formulae. $\xi(k)$ is **CKH***-consistent, so, by Theorem 8, there is a canonical k*-model of $\xi(k)$. The valuation φ is therefore defined so as to map each $k \in K$ to a canonical k*-model of $\xi(k)$.

Let then $\mathcal{M} = \langle k_0, K, \succ, \varphi \rangle$. It remains to be shown that $\mathcal{M}$ is a model of $\Gamma \cup \{\neg\psi\}$. More generally, we prove that for all $k \in \mathcal{M}$ and for every formula α, if $\alpha \in k$ then $\mathcal{M} \models_k \alpha$. The proof is by induction on the number of logical symbols in α.

1. If α is a k-formula and $\alpha \in k$, then $\alpha \in \xi(k)$. Therefore, by construction, $\varphi(k) \models \alpha$, i.e. $\mathcal{M} \models_k \alpha$.

2. If $\alpha = \neg\alpha_0$ and α_0 is not a k-formula, we distinguish different cases, according to the form of α_0.

   (a) $\alpha = \neg\neg\alpha_1$. If $\alpha \in k$ then $\alpha_1 \in k$, since k is a PC-expanded set. By induction hypothesis, $\mathcal{M} \models_k \alpha_1$, therefore $\mathcal{M} \models_k \alpha$.

   (b) $\alpha = \neg(\alpha_1 \vee \alpha_2)$. If $\alpha \in k$ then $\neg\alpha_1 \in k$ and $\neg\alpha_2 \in k$, since k is a propositional expansion. By induction hypothesis, $\mathcal{M} \models_k \neg\alpha_1$ and $\mathcal{M} \models_k \neg\alpha_2$, therefore $\mathcal{M} \models_k \neg(\alpha_1 \rightarrow \alpha_2)$.

   (c) $\alpha = \neg(\alpha_1 \wedge \alpha_2)$. If $\alpha \in k$ then either $\neg\alpha_1 \in k$ or $\neg\alpha_2 \in k$, since k is a propositional expansion. By induction hypothesis, either $\mathcal{M} \models_k \neg\alpha_1$ or $\mathcal{M} \models_k \neg\alpha_2$, therefore $\mathcal{M} \models_k \neg(\alpha_1 \rightarrow \alpha_2)$.

   (d) If $\alpha = \alpha_0 \rightarrow \alpha_1$ and $\alpha \in k$, then, either $\neg\alpha_0 \in k$ or $\alpha_1 \in k$, since k is a PC-expanded set. By induction hypotheses, in the first case $\mathcal{M} \models_k \neg\alpha_0$, in the second case $\mathcal{M} \models_k \alpha_1$. In both cases $\mathcal{M} \models_k \alpha$.

   (e) $\alpha = \neg[\star]\alpha_1$. If $\neg[\star]\alpha_1 \in k$, then, by construction of $\mathcal{M}$, there exists $k' \prec k$ such that $\neg\alpha_1 \in k'$. By induction hypothesis, $\mathcal{M} \models_{k'} \neg\alpha_1$, therefore $\mathcal{M} \models_k \neg[\star]\alpha_1$.

∎

The two previous lemmas entail the following:

Theorem 3 **AK4** *is sound and complete with respect to P-structures, i.e.* $\Gamma \models \alpha$ *iff* $\Gamma \vdash_{AK4} \alpha$*.*

4 The abductive relation

In this section, we assume that a given preference relation $<$ on formulae is given and we define the notion of P-models of $<$. We then show how the abductive relation, based on $<$, can be expressed by means of a P-formula. If the abductive relation holds between two PC-formulae A and B, the corresponding formula is true in every P-model of $<$, and vice-versa. This implies that, if Ω is a set of formulae axiomatizing $<$, i.e. such that for any P-structure $\mathcal{M}$, $\mathcal{M}$ is a model of Ω iff $\mathcal{M}$ is a model of $<$, then the abductive relation holds between two PC-formulae A and B iff the P-formula expressing this relation is derivable from Ω in **AK4**. In the next section some example of suche Ω's will be given.

The next definition introduces the basic semantical relation between preference structures and ordering relations on propositional formulae.

Definition 7 *Let* $<$ *be an ordering relation on PC-formulae. A preference structure* $\mathcal{M}$ *is a model of* $<$ *iff for all propositional formulae* A *and* B*, if* $\mathcal{M} \models \langle\star\rangle(\mathbf{O}A \wedge \langle\star\rangle\mathbf{O}B)$*, then* $B < A$*.*

In order for this definition to be meaningful, the existence of models of preference relations has to be ensured. This can be done only for orderings that satisfy the following property.

Definition 8 *An ordering relation* $<$ *on PC-formulae is* compatible with logical equivalence *iff for all PC-formulae* A, B, C, D *if* $\models_{PC} A \equiv B$ *and* $\models_{PC} C \equiv D$*, then* $A < C$ *iff* $B < D$*.*

Theorem 4 *For any ordering relation* $<$ *on PC-formulae, if* $<$ *is compatible with logical equivalence, then there is a model* $\mathcal{M}$ *of* $<$*.*

Proof. The model $\mathcal{M} = \langle k_0, K, \succ, \varphi\rangle$ is defined as follows. Let K^* be the set of the equivalence classes of PC-formulae: if $|A| = \{B \mid\ \models_{PC} A \equiv B\}$, then $K^* = \{|A| \mid A \text{ is a PC-formula}\}$, and let $K = K^* \cup k_0$, where k_0 is a new element. The function φ maps every $|A| \in K^*$ to the canonical k*-model of $\mathbf{O}A$, given by Lemma 2. Clearly, if $\models_{PC} A \equiv B$, then the canonical k*-model of $\mathbf{O}A$ is equal to the canonical k*-model of $\mathbf{O}B$. Therefore, for every PC-formula A, there is exactly one $k \in K^*$ such that $k \models \mathbf{O}A$. We establish that $\varphi(k_0)$ is any ok-model. The relation $\prec$ is defined by establishing that:

(a) for all $k \in K$, $k_0 \not\prec k$;
(b) for all $k \in K^*$, $k \prec k_0$;
(c) for all $|A|, |B| \in K^*$, $|A| \prec |B|$ iff $A < B$.

$\mathcal{M}$ is clearly a model of $<$. In fact, if $\mathcal{M} \models \langle\star\rangle(\mathbf{O}A \wedge \langle\star\rangle\mathbf{O}B)$ then there are elements $k_A, k_B \in K^*$ such that $\mathcal{M} \models_{k_A} \mathbf{O}A$, $\mathcal{M} \models_{k_B} \mathbf{O}B$, and $k_A \succ k_B$;

therefore $k_A = |A|$, $k_B = |B|$ and $B < A$. ∎

The P-model $\mathcal{M}$ built in the previous proof will be called the *canonical <-model*.

We can finally turn to the charaterization of the abductive relation. Let $\Theta \cup \{X\}$ be a set of PC-formulae and $\langle \Theta, X \rangle$ an abductive problem. We assume that Θ is a finite set of formulae, so, without loss of generality, we take Θ to be a single formula and we simply speak of explanations of formulae $\Theta \rightarrow X$, meaning that an explanation for the problem $\langle \Theta, X \rangle$ is a PC-formula A such that $\models_{PC} A \rightarrow (\Theta \rightarrow X)$. In the following any formula C such that $\models_{PC} C \rightarrow (\Theta \rightarrow X)$ is called an *admissible* explanation. Assuming that $<$ is fixed, the notion of minimal explanation is restated as follows.

Definition 9 *If A and B are PC-formulae, then A min-expl B iff $\models_{PC} A \rightarrow B$ and there is no $C < A$ such that $\models_{PC} C \rightarrow B$.*

The P-formula that we claim expresses the abductive relation is:

$$A \rightsquigarrow B \equiv_{def} [\star](\mathbf{O}A \rightarrow (\Box B \wedge [\star]\neg\Box B))$$

Intuitively, the formula states that: (i) $[\star](\mathbf{O}A \rightarrow \Box B)$, i.e. $\models_{PC} A \rightarrow B$; therefore, if $B = \Theta \rightarrow X$, A is an admissible explanation; (ii) $[\star](\mathbf{O}A \rightarrow [\star]\neg\Box B)$: for every canonical k*-structure k_C that may correspond to a formula C (i.e. $\mathcal{M} \models_{k_C} \mathbf{O}C$) that is better than A (i.e. $C < A$), $\mathcal{M} \models_{k_C} \mathbf{O}C \wedge \neg\Box B$ (i.e. $\not\models_{PC} C \rightarrow B$); therefore, C is not an admissible explanation.

The intuition is confirmed by the theorem that follows.

Theorem 5 *If A and B are PC-formulae, then A min-expl B iff for every model $\mathcal{M}$ of $<$, $\mathcal{M} \models A \rightsquigarrow B$.*

Proof. If A min-expl B, then:
(1) $\models_{PC} A \rightarrow B$ and
(2) for all PC-formulae C, it is not the case that $\models_{PC} C \rightarrow B$ and $C < A$.
Let $\mathcal{M}$ be a model of $<$ and let $k \in \mathcal{M}$ be such that $k \prec k_0$ and $\mathcal{M} \models_k \mathbf{O}A$; then $\mathcal{M} \models_k \Box B$ by (1) and Fact 3. Let k' be such that $k' \prec k$. Since every element in $\mathcal{M}$ is a canonical k*-structure, there is a formula C such that $\mathcal{M} \models_{k'} \mathbf{O}C$. Therefore $\mathcal{M} \models_k \mathbf{O}A \wedge \langle\star\rangle\mathbf{O}C$, so $\mathcal{M} \models_{k_0} \langle\star\rangle(\mathbf{O}A \wedge \langle\star\rangle\mathbf{O}C)$. Since $\mathcal{M}$ is a model of $<$, $C < A$. From (b) it follows that $\not\models_{PC} C \rightarrow B$, so, by Lemma 3, $A \not\models_{k'} \Box B$, i.e. $\mathcal{M} \models_{k'} \neg\Box B$; consequently $\mathcal{M} \models_k [\star]\neg\Box B$.

Conversely, let us assume that A min-expl B is false. We show that the canonical <-model $\mathcal{M}$, built in the proof of Theorem 4, is a countermodel of $A \rightsquigarrow B$. Let k_A be the unique $k \in \mathcal{M}$ such that $k \prec k_0$ and $\mathcal{M} \models_{k_A} \mathbf{O}A$. If A min-expl B is false because $\not\models_{PC} A \rightarrow B$, then $\mathcal{M} \not\models_{k_A} \Box B$ by Fact 3, so $\mathcal{M} \not\models [\star](\mathbf{O}A \rightarrow \Box B)$ and $\mathcal{M} \not\models A \rightsquigarrow B$. Let us now assume that there exists a formula $C < A$ such that $\models_{PC} C \rightarrow B$. Let $k_C \in \mathcal{M}$ be such that $k_C \prec k_0$ and $\mathcal{M} \models_{k_C} \mathbf{O}C$. By Fact 3, $\mathcal{M} \models_{k_C} \Box B$ and, by construction of the canonical model, $k_C \prec k_A$. Therefore $\mathcal{M} \not\models_{k_A} [\star]\neg\Box B$, so that $\mathcal{M} \not\models_{k_0} [\star](\mathbf{O}A \rightarrow [\star]\neg\Box B)$, i.e. $\mathcal{M} \not\models_{k_0} A \rightsquigarrow B$. ∎

5 The axiomatization of preference relations

Let us assume that Ω is a set of P-formulae axiomatizing a given preference relation $<$ among PC-formulae, i.e. for any P-structure $\mathcal{M}$, $\mathcal{M}$ is a model of Ω iff $\mathcal{M}$ is a model of $<$. Hence, by Theorem 5, if A and B are PC-formulae, then A min-expl B iff $\Omega \vdash_{AK4} A \rightsquigarrow B$. In this section we are giving two examples of Ω's axiomatizing commonly used preference relations.

Let us first consider the following preference relation:

$$A < B \quad \text{iff} \quad \models_{PC} B \rightarrow A \text{ and } \not\models_{PC} A \rightarrow B$$

and let Ω be the set of all the instances of the following schemata:

(IRR) $[\star](\mathbf{O}A \rightarrow [\star]\neg\mathbf{O}A)$

(IMP) $[\star](\overleftarrow{\Box}A \rightarrow [\star]\overleftarrow{\Box}A)$

The following result establishes that Ω is in fact an axiomatization of $<$.

Theorem 6 *For every P-structure $\mathcal{M}$, $\mathcal{M} \models \Omega$ iff $\mathcal{M}$ is a model of $<$.*

Proof. We first show that Ω is complete: let us assume that $\mathcal{M} \models \Omega$ and, for some A and B, $\mathcal{M} \models \langle\star\rangle(\mathbf{O}A \wedge \langle\star\rangle\mathbf{O}B)$. It follows that there exist $k, k' \in \mathcal{M}$ such that $\mathcal{M} \models_k \mathbf{O}A$, $\mathcal{M} \models_{k'} \mathbf{O}B$ and $k \succ k'$. Since $\mathcal{M} \models_k \mathbf{O}A$, $\mathcal{M} \models_k \overleftarrow{\Box}\neg A$ and, since $\mathcal{M}$ is a model of (IMP), $\mathcal{M} \models_k [\star]\overleftarrow{\Box}\neg A$, thus $\mathcal{M} \models_{k'} \overleftarrow{\Box}\neg A$. From this fact and $\mathcal{M} \models_{k'} \mathbf{O}B$, it follows that $\mathcal{M} \models_{k'} \overleftrightarrow{\Box}(\neg B \rightarrow \neg A)$, so that $\models_{PC} A \rightarrow B$. If it were the case that $\models_{PC} B \rightarrow A$, then $\models_{PC} A \equiv B$, so $\mathcal{M} \models_{k'} \mathbf{O}A$; therefore it would be $\mathcal{M} \models_k \mathbf{O}A \wedge \langle\star\rangle\mathbf{O}A$. But this is impossible if $\mathcal{M}$ is a model of (IRR).

Now we show that Ω is correct: let $\mathcal{M}$ be a model of $<$, i.e. for every A and B, if $\mathcal{M} \models \langle\star\rangle(\mathbf{O}A \wedge \langle\star\rangle\mathbf{O}B)$, then $B < A$. We prove that every instance of (IRR) is true in $\mathcal{M}$. Let us assume that for some PC-formula A and $k \in \mathcal{M}$, $\mathcal{M} \models_k \mathbf{O}A$, and let $k' \prec k$; if it were not the case that $\mathcal{M} \not\models_{k'} \neg\mathbf{O}A$, then it would be $\mathcal{M} \models_{k'} \mathbf{O}A$, so that $\mathcal{M} \models \langle\star\rangle(\mathbf{O}A \wedge \langle\star\rangle\mathbf{O}A)$; hence $\not\models_{PC} A \rightarrow A$, that is false.

In order to show that every instance of (IMP) is true in $\mathcal{M}$, let us take $k, k' \in \mathcal{M}$, with $k \succ k'$ and $\mathcal{M} \models_k \overleftarrow{\Box}A$. Since k' is a canonical k*-structure, for some B and C it is the case that $\mathcal{M} \models_k \mathbf{O}B$ $\mathcal{M} \models_{k'} \mathbf{O}C$; therefore

$\mathcal{M} \models_k \mathbf{O}B \wedge \langle\star\rangle\mathbf{O}C$, so that, since $\mathcal{M}$ is a model of $<$, $\models_{PC} B \rightarrow C$. From $\mathcal{M} \models_k \mathbf{O}B$ and $\mathcal{M} \models_k \overleftarrow{\Box}A$, it follows that $\models_{PC} \neg B \rightarrow \neg A$, i.e. $\models_{PC} A \rightarrow B$. Then $\models_{PC} A \rightarrow C$, i.e. $\models_{PC} \neg C \rightarrow \neg A$. Since $\mathcal{M} \models_{k'} \overleftarrow{\Box}\neg C$, $\mathcal{M} \models_{k'} \overleftarrow{\Box}\neg A$. ∎

Abduction based on the preference relation $<$ amounts to a quite trivial task: the abductive relation is in fact the equivalence relation, i.e. A min-expl B iff $\models_{PC} A \equiv B$. For example, if $B = (p \rightarrow q) \rightarrow (p \wedge q)$, B is logically equivalent to p, so p is the only abductive explanation of B. As a sample derivation, we cannot thus show but a derivation in **AK4** of $A \rightsquigarrow A$ from Ω. In the rightmost column, we briefly write "from x and y" when the corresponding formula derives from the lines x and y by means of a **K4**-valid inference.

1	$[\star](\mathbf{O}A \rightarrow \overleftarrow{\Box}\neg A)$	Definition of $\mathbf{O}A$
2	$[\star](\overleftarrow{\Box}\neg A \rightarrow [\star]\overleftarrow{\Box}\neg A)$	IMP
3	$[\star](\mathbf{O}A \rightarrow [\star]\overleftarrow{\Box}\neg A)$	from 1 and 2
4	$[\star](\mathbf{O}A \rightarrow [\star]\neg\mathbf{O}A)$	IRR
5	$[\star](\mathbf{O}A \rightarrow [\star](\neg\Box A \vee \overleftarrow{\Diamond}A))$	from 4
6	$[\star](\mathbf{O}A \rightarrow [\star]\neg\Box A)$	from 3 and 5
7	$[\star](\mathbf{O}A \rightarrow \Box A)$	Definition of $\mathbf{O}A$
8	$[\star](\mathbf{O}A \rightarrow (\Box A \wedge [\star]\neg\Box A))$	from 6 and 7

In order to rule out those trivial explanations for a given abductive problem that are equivalent to the problem itself, syntactical restriction are usually considered. A common one is that an abductive explanation has to be (equivalent to) a conjunction of literals. Let us call *C-formulae* those PC-formulae that are equivalent to conjunctions of literals.

Correspondingly, let us consider the following restriction of $<$:

$A \ll B$ iff $A < B$ and for no formula C it is the case that $\models_{PC} A \equiv B \vee C$

Thus, in particular, if $A \ll B$ and B is a C-formula, then A is a C-formula, too.

Let Ω^* consist of all the formulae in Ω plus the instances of

(CF) $[\star]\neg(\mathbf{O}A \wedge \langle\star\rangle\mathbf{O}(A \vee B))$

Ω^* is an axiomatization of $\ll$:

Theorem 7 *For every P-structure $\mathcal{M}$, $\mathcal{M} \models \Omega^*$ iff $\mathcal{M}$ is a model of $\ll$.*

Proof. Let us assume that $\mathcal{M} \models \Omega^*$ and that for some $k \in \mathcal{M}$, $\mathcal{M} \models_k \mathbf{O}A \wedge \langle\star\rangle\mathbf{O}B$. Since $\mathcal{M} \models \Omega$, $B < A$. Moreover, $\not\models_{PC} B \equiv A \vee C$, otherwise $\mathcal{M}$ would falsify the instance of (CF): $[\star]\neg(\mathbf{O}A \wedge \langle\star\rangle\mathbf{O}(A \vee C))$. Therefore $B \ll A$.

Conversely, let us assume that $\mathcal{M}$ is a model of $\ll$. Then $\mathcal{M}$ is also a model of $<$, so every instance of (IRR) and (IMP) is true in $\mathcal{M}$. Moreover, if for some formulae A and B and some $k \in \mathcal{M}$ it were the case that

$\mathcal{M} \not\models [\star]\neg(\mathbf{O}A \wedge \langle\star\rangle\mathbf{O}(A \vee B))$, then $\mathcal{M} \models \langle\star\rangle(\mathbf{O}A \wedge \langle\star\rangle\mathbf{O}(A \vee B))$, so it would be $A \vee B \ll A$, that is false. ∎

The main difference between abduction based on $<$ and $\ll$ is that, when $\ll$ is considered, $A \vee B$ is not better than A, so that A may be an abductive explanation of $A \vee B$. The following sample derivation shows that $\Omega^* \vdash_{AK4} p \rightsquigarrow (p \vee q)$. Here, the language $\mathbf{P}=\{p, q\}$.

1	$[\star](\mathbf{O}p \rightarrow \overleftarrow{\Box}\neg p)$	Definition of $\mathbf{O}p$
2	$[\star](\mathbf{O}p \rightarrow \Box p)$	Definition of $\mathbf{O}p$
3	$[\star](\mathbf{O}p \rightarrow \Box(p \vee q))$	from 2
4	$[\star]((\overleftarrow{\Box}\neg p \wedge \overleftarrow{\Diamond}q) \rightarrow \overleftarrow{\Diamond}(\neg p \wedge q))$	valid, by K' and Nec'
5	$[\star](\Box\neg(\neg p \wedge q) \vee \overleftarrow{\Box}\neg(\neg p \wedge q))$	FIN
6	$[\star](\overleftarrow{\Diamond}(\neg p \wedge q) \rightarrow \Box\neg(\neg p \wedge q))$	from 5
7	$[\star]((\overleftarrow{\Box}\neg p \wedge \overleftarrow{\Diamond}q) \rightarrow \Box\neg(\neg p \wedge q))$	from 4 and 6
8	$[\star]((\overleftarrow{\Box}\neg p \wedge \overleftarrow{\Diamond}q \wedge \Box(p \vee q)) \rightarrow (\Box(p \vee \neg q) \wedge \Box(p \vee q)))$	from 7
9	$[\star]((\overleftarrow{\Box}\neg p \wedge \overleftarrow{\Diamond}q \wedge \Box(p \vee q)) \rightarrow \Box p)$	from 8
10	$[\star]((\overleftarrow{\Box}\neg p \wedge \overleftarrow{\Diamond}q \wedge \Box(p \vee q)) \rightarrow \mathbf{O}p)$	from 9
11	$[\star](\mathbf{O}p \rightarrow [\star]\neg\mathbf{O}p)$	IRR
12	$[\star](\overleftarrow{\Box}\neg q \wedge \overleftarrow{\Box}\neg p \rightarrow \overleftarrow{\Box}\neg(p \vee q))$	valid, by K' and Nec'
13	$[\star](\overleftarrow{\Box}\neg q \wedge \overleftarrow{\Box}\neg p \wedge \Box(p \vee q) \rightarrow \mathbf{O}(p \vee q))$	from 12
14	$[\star]\neg(\mathbf{O}p \wedge \langle\star\rangle\mathbf{O}(p \vee q))$	CF
15	$[\star](\mathbf{O}p \rightarrow [\star]\neg\mathbf{O}(p \vee q))$	from 14
16	$[\star](\mathbf{O}p \rightarrow [\star](\overleftarrow{\Box}\neg p \wedge \neg\mathbf{O}p \wedge \neg\mathbf{O}(p \vee q)))$	from 1, 11 and 15
17	$[\star](\mathbf{O}p \rightarrow [\star](\neg(\overleftarrow{\Diamond}q \wedge \Box(p \vee q)) \wedge \neg(\overleftarrow{\Box}\neg q \wedge \Box(p \vee q))))$	from 10, 13 and 16
18	$[\star](\mathbf{O}p \rightarrow [\star]\neg\Box(p \vee q))$	from 17 by PC
19	$[\star](\mathbf{O}p \rightarrow (\Box(p \vee q) \wedge [\star]\neg\Box(p \vee q)))$	from 3 and 18

6 Concluding remarks

In this work, a logic for the abductive relation has been defined, whose semantics relies on the models of a preference relation $<$. The formulations of abduction that are based upon the classical approach can be embedded in this semantic framework, by means of a suitable characterization of the underlying preference relation. The fact that this logic turns out to be quite cumbersome should not surprise, due to the strong interplay of different concepts in the meta-level definition of abduction, such as logical validity and minimality in a set of formulae.

By means of the semantics and proof theory of the logic, structural properties of the abductive relation can be proved, that are independent from any particular preference relation. For example, the fact that $A \rightsquigarrow B, A \rightsquigarrow C \vdash_{AK4} A \rightsquigarrow (B \wedge C)$ shows that the abductive implication enjoys the "and-right" property:

1.	$\mathbf{O}A \rightarrow (\Box B \wedge [\star]\neg\Box B)$	Assumption
2.	$\mathbf{O}A \rightarrow (\Box C \wedge [\star]\neg\Box C)$	Assumption
3.	$\mathbf{O}A \rightarrow (\Box B \wedge \Box C)$	From 1,2, by PC
4.	$(\Box B \wedge \Box C) \rightarrow \Box(B \wedge C)$	Theorem of KH
5.	$\mathbf{O}A \rightarrow \Box(B \wedge C)$	From 3,4, by PC
6.	$\neg\Box B \rightarrow \neg\Box(B \wedge C)$	Theorem of KH
7.	$[\star](\neg\Box B \rightarrow \neg\Box(B \wedge C))$	From 6, by $[\star] - Nec$
8.	$[\star](\neg\Box B) \rightarrow [\star](\neg\Box(B \wedge C))$	From 7
9.	$\mathbf{O}A \rightarrow [\star]\neg\Box(B \wedge C))$	From 1 and 8, by PC
10.	$\mathbf{O}A \rightarrow (\Box(B \wedge C) \wedge [\star]\neg\Box(B \wedge C))$	From 5 and 9, by PC

Conversely, the fact that some properties do not hold in general can be proved semantically, by exhibiting a countermodel. For example, monotonicity and right weakening do not hold: in fact, $A \leadsto B \not\models_{AK4} A \wedge C \leadsto B$ and $A \leadsto B \not\models_{AK4} A \leadsto B \vee C$.

The logic that has been defined aims at capturing the notion of abductive relation. The study is still very preliminary and leaves a number of open problems, such as the extensions to an infinite language and/or infinite background theory, the possibility of (partially) authomatizing deduction, the axiomatization of other interesting preference relations, the extension to different logics and, in particular, to non monotonic logics.

A Weak compelteness of KH* over an infinite language

In this appendix whe show that the system **KH*** is weakly complete with respect to canonical K*-structures, whenever the the language **P**$= \{p_0, p_1, ...\}$ is infinite. The result is established along the lines of [Lev90]. Here the restriction on the application of modal operators to propositional formulae is effectively used. In fact, contrarily to the system considered in [Lev90], based upon modal S5, in **KH*** modal functions cannot be reduced.

We shall use the following notational convention. If $k = \langle w_0, W, R, V\rangle$ is a k*-structure, then:

$$W_k^+ = \{w \in W \mid w_0 R w\} \qquad W_k^- = \{w \in W \mid w_0 \overline{R} w\}$$

When it is clear from the context, the subscript k will be omitted.

As usual, *literal* is either a propositional letter or the negation of a propositional letter. A *modal literal* is a k-formula of either the form $\star A$ or $\neg\star A$, where $\star$ is either $\Box$ or $\overleftarrow{\Box}$.

Lemma A *Let A and B be PC-fromulae. If $\models \mathbf{O}B \rightarrow A$, then $\models_{PC} A$.*

Proof. We show that if $\not\models_{PC} A$, then $\not\models \mathbf{O}B \rightarrow A$. Let ass be an assigment such that $ass(A) = false$ and let $k = \langle w_0, W, R, V\rangle$ be the canonical k*-model of $\mathbf{O}B$. We define the k*-structure $k' = \langle w_0', W', R', V'\rangle$ as follows. $W' = W$, $V' = V$. w_0' is the element of W such that $V(w_0') = ass$. The relation R' is defined as follows:

for every $w, w' \in W$, $wR'w'$ iff $w = w_0'$ and w_0Rw'

We show that $k' \models_{w_0'} \mathbf{O}B$. Let $w \in W$. If $w_0'R'w$, then w_0Rw, therefore $V(w)(B) = true$, so that $k' \models_{w_0'} \Box B$. If $w_0'\overline{R'}w$, then $w_0\overline{R}w$, therefore $V(w)(B) = false$, so that $k' \models_{w_0'} \overleftarrow{\Box}\neg B$. Finally, $k' \not\models_{w_0'} A$, because $V(w_0')(A) = ass(A) = false$. Therefore k' is a k*-structure falsifying $\mathbf{O}B \rightarrow A$. ∎

The following lemma is lemma A.2.1. in [Lev90]. The proof is essentially the same.

Lemma B *If A and B are PC-formulae such that $A \wedge \neg B$ is satisfiable, then:*

(a) $\vdash_{KH^*} \Box\neg A \rightarrow \neg\overleftarrow{\Box}B$, therefore $\Box\neg A \rightarrow \neg\overleftarrow{\Box}B$ is valid.

(b) $\vdash_{KH^*} \overleftarrow{\Box}\neg A \rightarrow \neg\Box B$, therefore $\overleftarrow{\Box}\neg A \rightarrow \neg\Box B$ is valid.

Proof.

(a) The following is a proof of $\Box\neg A \rightarrow \neg\overleftarrow{\Box}B$ in **KH***:

1.	$\overleftrightarrow{\Diamond}(A \wedge \neg B)$	by LP, since $A \wedge \neg B$ is satisfiable
2.	$\Box\neg(A \wedge \neg B) \rightarrow \overleftarrow{\Diamond}(A \wedge \neg B)$	from 1, by definitions and PC
3.	$\Box\neg A \rightarrow \Box\neg(A \wedge \neg B)$	K-theorem
4.	$\overleftarrow{\Diamond}(A \wedge \neg B) \rightarrow \neg\overleftarrow{\Box}B$	K'-theorem
5.	$\Box\neg A \rightarrow \neg\overleftarrow{\Box}B$	from 2, 3, 4, by PC

A direct semantical proof can also be given: let k be any k*-structure. If $A \wedge \neg B$ is satisfiable, then there is some $w \in k$ such that $k \models_w A \wedge \neg B$. If $k \models_{w_0} \Box\neg A$, where w_0 is the actual world in k, then such a $w \in k$ cannot belong to W^+, but $w \in W^-$. As $k \models_w \neg B$, $k \models_{w_0} \neg\overleftarrow{\Box}B$.

(b) The following is a proof of $\Box\neg A \rightarrow \neg\overleftarrow{\Box}B$ in **KH***:

1.	$\overleftrightarrow{\Diamond}(A \wedge \neg B)$	by LP
2.	$\overleftarrow{\Box}\neg(A \wedge \neg B) \rightarrow \Diamond(A \wedge \neg B)$	from 1, by definitions and PC
3.	$\overleftarrow{\Box}\neg A \rightarrow \overleftarrow{\Box}\neg(A \wedge \neg B)$	K'-theorem
4.	$\Diamond(A \wedge \neg B) \rightarrow \neg\Box B$	K-theorem
5.	$\overleftarrow{\Box}\neg A \rightarrow \neg\Box B$	from 2, 3, 4, by PC

A direct semantical proof can also be given: let k be any k*-structure. If $A \wedge \neg B$ is satisfiable, then there is some $w \in k$ such that $k \models_w A \wedge \neg B$. If $k \models_{w_0} \overleftarrow{\Box}\neg A$, where w_0 is the actual world in k, $w \in W^+$. As $k \models_w \neg B$, $k \models_{w_0} \neg\Box B$.

∎

Lemma C *If A and C are PC-formulae such that $\Box A \wedge \overleftarrow{\Box} B$ is consistent, then* $\models_{PC} A \vee B$

Proof. If $\Box A \wedge \overleftarrow{\Box} B$ is consistent, then there exists a k*-structure $k \models \Box A \wedge \overleftarrow{\Box} B$. By K- and K'-transformations, $k \models \Box(A \vee B) \wedge \overleftarrow{\Box}(A \vee B)$, i.e. $k \models \overleftrightarrow{\Box}(A \vee B)$. The conclusion follows from the fact that k is a k*-structure.
The statement can also be proved by use of H. In fact, $(\Box A \wedge \overleftarrow{\Box} B) \rightarrow \overleftrightarrow{\Box}(A \vee B)$ is derivable. If $\Box A \wedge \overleftarrow{\Box} B$ is consistent, then there exists a k*-structure $k \models \Box A \wedge \overleftarrow{\Box} B$, so $k \models \overleftrightarrow{\Box}(A \vee B)$. The conclusion follows as before. ∎

The following lemma is a restatement of Lemma A.2.5 in [Lev90]. The proof is essentially the same.

Lemma D *If σ is a consistent conjunction of modal literals, and if there is some propositional letter in the language that does not occur in σ, then there is a PC-formula E such that $\mathbf{O}E \rightarrow \sigma$ is valid.*

Proof. Let A, B, C, D be finite sets of PC-formulae, and let

$$A^* = \{\Box A_i \mid A_i \in A\} \qquad B^* = \{\neg\Box B_i \mid B_i \in B\}$$
$$C^* = \{\overleftarrow{\Box} C_i \mid C_i \in C\} \qquad D^* = \{\neg\overleftarrow{\Box} D_i \mid D_i \in D\}$$

We may assume that σ is the conjunction of the literals in the union $\Sigma = A^* \cup B^* \cup C^* \cup D^*$. Obviously, some of A^*, B^*, C^*, D^* may be empty.

Let p be any propostional letter that does not occur in σ and let

$$E = \neg C \vee (p \wedge A)$$

If either C or A is empty, then the respective part of E is replaced by $\top$.

We show that, for all $X \in \Sigma$, $\mathbf{O}E \rightarrow X$ is valid. In order to do that, let us assume $k \models \mathbf{O}E$, i.e. for all $w \in k$, if $w \in W^+$, then $k \models_w \neg C \vee (p \wedge A)$, and if $w \in W^-$, then $k \models_w C \wedge \neg(p \wedge A)$. We show that, for all $X \in \Sigma$, $k \models X$.

(A) Let $X = \Box A_i$. If σ is consistent, then in particular, $\Box A_i \wedge C^*$ is consistent, hence, by K'-transformations, $\Box A_i \wedge \overleftarrow{\Box} C$ is consistent. If $C \neq \emptyset$, then, by lemma C, $\models_{PC} (A_i \vee C)$, i.e. $\models_{PC} \neg C \rightarrow A_i$. Obviously, $\models_{PC} A \rightarrow A_i$, consequently, $\models_{PC} E \rightarrow A_i$. If $C = \emptyset$, then $E = p \wedge A$, so, again, $\models_{PC} E \rightarrow A_i$. Since $k \models \Box E$, it follows that $k \models \Box A_i$.

(B) Let $X = \neg\Box B_i$. If σ is consistent, then so is $A^* \wedge \neg\Box B_i$, and, by K-transformations, $\Box A \wedge \neg\Box B_i$ is consistent. Therefore $A \wedge \neg B_i$ is consistent and, since p does not occur in σ, $(p \wedge A) \wedge \neg B_i$ is consistent. By Lemma B.b, $\overleftarrow{\Box}\neg(p \wedge A) \rightarrow \neg\Box B_i$ is valid. By hypothesis, $k \models \overleftarrow{\Box}\neg(p \wedge A)$, therefore $k \models \neg\Box B_i$. Note that the above reasoning holds also when A is empty, thanks to the introduction of the new letter p.

(C) If $X = \overleftarrow{\Box}C_i$, then, since $k \models \overleftarrow{\Box}C$, $k \models \overleftarrow{\Box}C_i$.

(D) Let $X = \neg\overleftarrow{\Box}D_i$. Since σ is consistent, so is $C^* \wedge \neg\overleftarrow{\Box}D_i$ and, by K'-transformations, $\overleftarrow{\Box}C \wedge \neg\overleftarrow{\Box}D_i$ is consistent. Therefore, $C \wedge \neg D_i$ is propositionally consistent and so is $(C \wedge \neg p) \wedge \neg D_i$. By Lemma B.a, $\Box\neg(C \wedge \neg p) \rightarrow \neg\overleftarrow{\Box}D_i$, i.e. $\Box(\neg C \vee p) \rightarrow \neg\overleftarrow{\Box}D_i$, is valid. By hypothesis, $k \models \Box(\neg C \vee p)$, so $k \models \neg\overleftarrow{\Box}D_i$. Note again that the above reasoning holds also when C is empty, thanks to the introduction of the new letter p.

∎

The following corresponds to Lemma A.2.9 in [Lev90]. The proof is essentially the same.

Lemma 9 *Let α be a k-formula such that there is some propositional letter in the language that does not occur in it. If for every PC-formula E, $\models \mathbf{O}E \rightarrow \alpha$, then $\models \alpha$.*

Proof. Every k-formula is built up from literals and modal literals by means of propositional operators. By means of propositional transformations, α can be reduced into an equivalent conjunctive normal form, where the literals are either propositional or modal. Let us assume α is the conjunction of formulae of the form $A_i \vee \sigma_i$, where the A_i are disjunctions of propositional literals and the σ_i are disjunctions of modal literals. We show that every such $A_i \vee \sigma_i$ is valid.

Let us assume first that σ_i is not missing and that it is not valid. Therefore $\neg\sigma_i$ is consistent and, since it is a conjunction of modal literals, and there is some propositional letter in the language that does not occur in it, by Lemma D, there exists a propositional formula E such that (i) $\mathbf{O}E \rightarrow \neg\sigma_i$ is valid. By hypothesis, $\models \mathbf{O}E \rightarrow \alpha$, then also (ii) $\models \mathbf{O}E \rightarrow (A_i \vee \sigma_i)$. If A_i is present, then from (i) and (ii) it follows that $\mathbf{O}E \rightarrow A_i$ is valid. By Lemma A, A_i is valid, therefore $A_i \vee \sigma_i$ is valid. If, on the contrary, A_i is missing, then by (i) and (ii) it follows that $\mathbf{O}E$ is inconsistent, contradicting Lemma 2. Therefore the hypothesis that σ_i is not valid is absurd.

If σ_i is missing, then, since by hypothesys, $\mathbf{O}p \rightarrow A_i$, A_i is valid by Lemma A. ∎

Lemma 10 *If α is a satisfiable k-formula, then there exists a PC-formula A sucht that $\alpha \wedge \mathbf{O}A$ is satisfiable.*

The following form of weak completeness is therefore a straightforward consequence of the completeness of **KH*** with respect to k*-structures.

Theorem 8 *If α is a k-formula, then $\vdash_{KH^*} \alpha$ iff α is true in every canonical structure.*

Proof. Soundness follows directly from the soundness of **KH*** with respect to k*-structures. Completeness is shown as follows. If α is true in every canonical structure, then for every PC-formula E, $\models \mathbf{O}E \rightarrow \alpha$. Therefore α is valid by Lemma 9, and $\vdash_{KH*} \alpha$ follows from the completeness of **KH*** with respect to k*-structures. ■

Strong completeness does not hold: there exist infinite sets Γ of k-formulae that are satisfiable in some k*-structure, but have no canonical model.

Lemma E *There exist infinite sets of k-formulae that are satisfiable but do not have any canonical model.*

Proof. An example is given by $\Gamma = \{\neg\mathbf{O}A \mid A \in \mathcal{F}_{PC}\}$, where $\mathcal{F}_{PC}$ is the set of all the PC-formulae in the language. Clearly, Γ has no canonical model. However, the k*-structure $k = \langle w_0, W, R, V\rangle$, defined as follows, is a model of Γ. If $|\mathbf{S}|$ is the cardinality the set $\mathbf{S}$ of all the possible interpretations of $\mathbf{P}$, then W is a set having $1 + |\mathbf{S}|$ elements. W^- is a singleton $\{w^-\}$ and W^+ has $|\mathbf{S}|$ elements. V is obtained by extending any one-to-one mapping from W^+ onto $\mathbf{S}$ to W, with $V(w^-) = ass$, where ass is any element of $\mathbf{S}$. If A is a PC-formula such that $k \models \Box A$, then A is a tautology. But, as W^- is non empty, $k \not\models \overleftarrow{\Box}\neg A$. ■

References

[BB93a] C. Boutillier and V. Becher. Abduction as belief revision. In *Proceedings of the Eleventh National Conference on Artificial Intelligence (AAAI-93)*, pages 642–648, 1993.

[BB93b] C. Boutillier and V. Becher. Abduction as belief revision. Technical Report 93-23, University of British Columbia, 1993.

[Bou94] C Boutilier. Conditional logics of normality: a modal approach. *Artificial Intelligence Journal*, 68:87–154, 1994.

[CMP93] M. Cialdea Mayer and F. Pirri. First order abduction via tableau and sequent calculi. *Bulletin of the IGPL*, 1:99–117, 1993.

[CMP95] M. Cialdea Mayer and F. Pirri. Modal propositional abduction. *Bulletin of the IGPL*, 6(3):99–117, 1995.

[CP86] P. T. Cox and T. Pietrzykowski. Causes for events: their computation and applications. In *Proceedings of the Eighth Conference on Automated Deduction (CADE-86)*, pages 608–621, 1986.

[DFndC91] R. Demolombe and L. Fariñas del Cerro. An inference rule for hypothesis generation. In *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence (IJCAI-91)*, pages 152–157, 1991.

[G$\ddot{9}$0] P. Gärdenfors. An epistemic analysis of Explanations and Causal beliefs. *Topoi*, 9:109–124, 1990.

[GP92] V. Goranko and S. Passy. Using the universal modality: gains and questions. *Journal of Logic and Computation*, 2:5–29, 1992.

[GPT86] Gargov, Passy, and Tinchev. Modal enviroment for boolean speculations. In *Proceedings of the 1986 Goedel Conference in Bulgaria*, pages 253–263, 1986.

[Har78] G.H. Harman. The inference to the best explanation. *Philosophical Review*, 74:88–95, 1978.

[Hem62] C.G. Hempel. Deductive-nomological vs. statistical explanation. *Minnesota Studies in the Philosophy of Science*, 3:98–169, 1962.

[Hum83] I.L. Humberstone. Inaccessible worlds. *Notre Dame Journal of Formal Logic*, 24:346–352, 1983.

[Hum87] I.L. Humberstone. The modal logic of 'All and Only'. *Notre Dame Journal of Formal Logic*, 28:177–188, 1987.

[Lev89] H. J. Levesque. A knowledge-level account of abduction. In *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence (IJCAI-89)*, pages 1061–1067, 1989.

[Lev90] H. J. Levesque. All I know: a study in autoepistemic logic. *Artificial Intelligence Journal*, 42:381–386, 1990.

[Pei62] C.S. Peirce. Abduction and induction. In P. Owen, editor, *Classics in Logic*, pages 331–352. Academic Press, 1962.

[Poo88] D. Poole. Representing knowledge for logic-based diagnosis. In *Proceedings of the International Conference on Fifth Generation Computer Systems*, pages 1282–1289, 1988.

[RdK87] R. Reiter and J. de Kleer. Foundations of assumption-based truth maintenance systems: Preliminary report. In *Proceedings of the National Conference on Artificial Intelligence*, pages 183–188, 1987.

[Tha78] P. Thagard. The best explanation: criteria for theory choice. *Journal of Philosophy*, 75:76–92, 1978.