Counterfactual Reasoning Based on Revising Assumptions

Luís Moniz Pereira Joaquim Nunes Aparício
José Júlio Alferes
AI Centre, Uninova, 2825 Monte da Caparica, Portugal
({lmp,jna,jja}@fct.unl.pt)

Abstract

We present a semantics for counterfactual implication relative to positive programs with integrity constraints and protected rules (unrevisable). A counterfactual truth evaluation is presented based on *Contradiction Removal Semantics*. We show the adequacy of CRS to define the notion of similarity needed for defining counterfactual truth. The concepts introduced by CRS are adequate to support the revision process required by contravening hypoteses when evaluating counterfactuals.

1 Introduction

We present a semantics for counterfactual implication relative to positive logic programs with integrity constraints and protected (unrevisable) rules.The Contradiction Removal Semantics (CRS) [PAA91a] setting employed is a suitable extension to well founded model semantics (WFS) [VGRS90], to deal with classical negation and contradiction, by allowing the revising of assumptions. Lewis [Lew73] is a classic essay on counterfactuals, and our basic definitions and concepts drawn on it. In the next section CRS semantics is reviewed. In the third, Lewis's classic proposal for counterfactuals is introduced. We assume here the reader is familiar with WF Semantics as presented, say, in [VGRS90, Prz90, PAA91b]. In what follows, we consider a program as being a set of rules of the form:

$$H \leftarrow B_1, B_2, \ldots, B_n, \sim C_1, \sim C_2, \ldots, \sim C_m \quad (m \geq 0, n \geq 0)$$

The symbol $\sim$stands for negation by default, i.e. negation in the sense of WFS. $H, B_1, B_2, \ldots, B_n, C_1, C_2$, and C_m are literals. A predicate symbol is defined as usual. A negative predicate symbol is a predicate symbol whose first character is the special symbol $\neg$. The intended meaning of a negative literal $\neg L$ is the classical negation of positive literal L. As in [GL90] and in [Prz90], we transform these programs into ones obtained by replacing every occurrence of every negated classical literal $\neg p$ with a negative predicate symbol $\neg p$[1].Their approach has however a problem, in the sense that often we end up with no models at all, thus failing to provide a semantics to a (contradictory) program. This occurs because models containing both the positive and negative literals are discarded in this approach. We argue that this is too strong, and that a further refinement of the semantics can provide

[1]Actually they suggest replacing every occurrence of a literal $\neg p$ with atom p'.

intuitive and desired results in the presence of contradictory information. Dealing with contradiction in a constructive way is of course essential for counterfactual reasoning.

2 Counterfactuals

Counterfactuals are usually defined based on a notion of possible (or alternative) worlds (or simply worlds), and the definition of the truth value of a counterfactual $p > q$ considers the possible worlds "maximally similar" to the world S, where the counterfactual is being evaluated. We don't need to quantify how similar two worlds S_1 and S_2 are, but only to define a partial order between worlds, enabling to determine whether S_1 is more similar than S_2 with respect to S .

2.1 Lewis's counterfactuals

Given a similarity relation $\geq_F$ that nests layers of similar worlds relative to F , obeying some reasonable precepts, Lewis [Lew73] defines a counterfactual $p > q$ to be true in some world F only in either of two cases:

- (A) if there exists some world U such that:
  - (A1) p is true in U and
  - (A2) $p \rightarrow q$ is true in any world V with $V \geq_F U$ (i.e. if V is any world at least as similar to F as U is, q will hold in V if p does);
- (B) if there is no world U in which p holds $p > q$ is vacuously true.

Lewis uses $\Box\!\!\rightarrow$ in place of $>$. Let p-world mean a world where p holds. He also provides a "might" counterfactual defined as $p \Diamond\!\!\rightarrow q =_{def} \neg(p \Box \rightarrow \neg q)$, meaning "there is a nearest p-world where $p \rightarrow q$".

Counterfactual implication is also equivalently defined by Lewis [Lew73] in a different way, which we adopt, directly based on a notion of maximally similar worlds. Let $f(p, w)$ stand for the set of nearest p-worlds from w.

$$p >_w q =_{def} f(p, w) \subset [\![q]\!]$$

where f is a selection function [Lew73, page 57], where $[\![q]\!]$ represents the set of worlds in which q is true, and meaning: "the counterfactual implication $p > q$ is true in world w iff in all the nearest p-worlds from w, q is also true"; in other words, if the set of the nearest p-worlds is a subset of the q-worlds (if the set of nearest p-worlds is empty then $p > q$ is vacuously true).

Definition 2.1 (Selection function) *([Lew73])* A function f from sentences and worlds to sets of worlds is a *selection function* if and only if, for all sentences ϕ and ψ and for each world i, the following four conditions hold.

i) If ϕ is true at i then $f(\phi, i)$ is the set $\{i\}$ having i as its only member.

ii) $f(\phi, i)$ is included in $[\![\phi]\!]$.

iii) If $[\![\phi]\!]$ is included in $[\![\psi]\!]$ and $f(\phi, i)$ is nonempty then $f(\psi, i)$ is also nonempty.

iv) If $[\![\phi]\!]$ is included in $[\![\psi]\!]$ and $[\![\phi]\!]$ overlaps $f(\psi, i)$ then $f(\phi, i)$ is the intersection of $[\![\phi]\!]$ and $f(\psi, i)$.

3 Worlds as Models

In this section we interpret a world as a set of literals, which is a model of a given program or a subset of that model. Worlds may be contradictory, in the sense that a literal L and the literal which stands for its classical negation $\neg L$ are both present in the model.

Consider the following program: $P = \{f(X) \leftarrow b(X), bf(X)[1] \quad bf(X) \leftarrow \sim \neg bf(X)\ [2] \quad b(a)\ [3] \quad b(b)\ [4] \quad \neg f(b)\ [5]\}$, where rule [1] says that birds fly if the rule is applicable, and rule [2] says that you may assume (by closed world assumption on $\neg bf(X)$) that rule [1] is applicable, i.e. $bf(c)$ holds for some c, if there is no evidence that the rule is not applicable, (i.e. $\neg bf(c)$ holds). Thus, use of rule [1] when $X = c$ depends on the CWA-"assumption" $\sim\neg bf(c)$.

The well founded model $WFM(P) \models \{\neg f(b), f(b), \ldots\}$, reveals a contradiction. There is however no contradiction concerning object a. The complete $WFM(P)$ is

$$\{f(a), b(a), bf(a), \sim\neg bf(a), f(b), b(b), bf(b), \sim\neg bf(b), \neg f(b)\}.$$

We show how contradictory worlds (e.g. contradictory well founded models) can be revised by first questioning contradictory conclusions supported by assumptions made so far (e.g. $\sim\neg bf(b)$ above).

The contradiction is supported by "assumption" $\sim\neg bf(b)$, which belongs to the well founded model. We sustain that from this P we still should be able to derive all of $\{f(a), b(a), bf(a), \sim\neg bf(a), b(b), \neg f(b)\}$.

We will use the concepts introduced in CRS [PAA91a] to provide a semantics for a program P with a contradictory $WFM(P)$, and show through examples that it provides a contradiction revision method, adequate to support the revision process required to deal with contravening hypoteses [Res64] of counterfactuals.

4 Contradiction Removal Semantics (CRS)

In this section we review some definitions of CRS, taken from [PAA91a].

Definition 4.1 (Support set of a literal) A Support Set of a literal L belonging to the WF Model, M_P, of a program P, represented as $SS_P(L)$, or $SS(L)$ for short, is obtained as follows:

- If L is an atom, choose some rule of P for L where all the literals in its body belong to M_P. One $SS(L)$ is obtained by taking all those body literals plus the literals in some SS of each body literal.

- If $L = \sim A$, if there are no rules defined for A in P then the only SS of L is $\{\}$; otherwise, choose from each rule defined for A, a literal such that its complement belongs to M_P. A $SS(L)$ has all those complement literals, and the literals of a SS of each of them.

By considering all possible rules of P for some literal all its SSs are obtained.

Definition 4.2 (Assumption support set of a literal) An Assumption Support Set (ASS) of a literal L is the (possibly empty) subset of some support set of L, which contains all its $\sim$negative elements having no rules for their complement. We represent the ASS of a literal L in a program P as: $ASS_P(L)$ or $ASS(L)$ for short.

Definition 4.3 (Removal set of a literal) A Removal Set (RS) of a literal L belonging to the WFM of program P, is a set of atoms built from the union of some nonempty subset from each $ASS_P(L)$. If the empty set is an $ASS_P(L)$ then the only $RS(L)$ is the empty set. Note that a literal not belonging to the WFM of P has no RSs defined.

Definition 4.4 (Contradiction removal set of a program) A Contradiction Removal Set (CRS) of a program P is a Minimal Removal Set of the (reserved) literal $\perp$, i.e. a CRS of P is a $MRSP(\perp)$.

Definition 4.5 (Neighbour contradiction free program) A Neighbour Contradiction Free Program ($NCFP$) of a program P with CRS C_i, is the program obtained by adding to P a rule $L \leftarrow \sim L$ (called a CWA-inhibition rule in [PAA91a]) for each $L \in C_i$.

Note that a program with $n \geq 0$ CRSs can be transformed into n different $NCFP$s. If a program P has no CRSs then the only $NCFP$ is defined as P itself.

Definition 4.6 (Submodel of a program) Consider a contradictory program P whose WFM has negative literals $L_1, L_2, \ldots, L_n$ true by Closed World Assumption CWA. Let $Lits_{CWA} = \{L_1, L_2, \ldots, L_n\}$. A submodel of P is the $WFM$of the program obtained by adding CWA inhibition rules for each element of some subset of $Lits_{CWA}$, thereby making them undefined in the WFM of the derived program.

5 Counterfactual Reasoning by Revising Assumptions

Definition 5.1 (Program Base) A Program Base $PB = \langle K; S; IC \rangle$ is a set of rules K of the form, $H \leftarrow B_1, \ldots, B_n (n \geq 0)$ where B_i are atoms[2] and S is a set of protected rules, with $S \subseteq K$, and a set IC of integrity constraints.

We denote without loss of generality an integrity constraint as a rule $\perp \leftarrow L_1, \ldots, L_n$, where $\perp$ is a reserved symbol and L_i's are atoms. By the definition above the set $K \cup IC$ is always a positive logic program.

The semantics of a Program Base is defined by the least model [VEK76] $M = least(K \cup IC)$ if it is non-contradictory.

Definition 5.2 (Contradictory PB) A Program Base PB is contradictory iff $\perp \in least(K \cup IC)$.

Definition 5.3 (Program associated with a Program Base) The program P associated with a Program Base $PB = \langle K; S; IC \rangle$ is a logic program obtained from PB as follows:

- for each unprotected rule $H \leftarrow B_1, \ldots, B_n$ $(n \geq 0)$ in K, P contains instead the rule $H \leftarrow B_1, \ldots, B_n, \sim A$ where A is a new atom.
- P contains all the other rules in K and IC.

[2] This definition is not more general, i.e. the B_i's could be negative literals, because it can be shown that the properties Lewis's original selection function must obey are only guaranteed for this type of Program Base. A new type of selection function is needed to account for the presence of $\sim$.

This defines a transformation $\mathcal{T} : PB \rightarrow P$ from program bases to logic programs.

Example 1 Consider the Program Base $PB_1 = \langle\{a \leftarrow b\};\{\};\{\}\rangle$. Its associated program is $P_1 = \{a \leftarrow b, \sim c\}$. We have $least(PB) = \{\}$ and $WFM(P_1) = \{\sim a, \sim b, \sim c\}$. Consider the Program Base $PB_2 = \langle\{b, a \leftarrow b\};\{b\};\{\}\rangle$. Its associated program is $P_2 = \{b, a \leftarrow b, \sim c\}$. We have $least(PB_2) = \{a, b\}$ and $WFM(P_2) = \{a, b, \sim c\}$.

Definition 5.4 (3-valued interpretation) *([Prz89])* A 3-valued Herbrand interpretation I of a language L is a pair $\langle T; F\rangle$, where T and F are disjoint subsets of the Herbrand base B. T contains all ground atoms **true** in I , F contains all ground atoms **false** in I , and the truth value of the remaining atoms, those in $U = B - (T \cup F)$, is **undefined**.

We also represent an interpretation $I = \langle T; F\rangle$ as $I = T \cup \{\sim L | L \in F\}$.

Definition 5.5 (Contradictory model) A contradictory model M is a model such that $\perp \in M$.

Example 2 Consider the "flying birds (and penguins)" Program Base, with $K = \{f \leftarrow b\ [1)\quad \neg f \leftarrow p[2)\quad \perp \leftarrow f, \neg f[3]\quad b[4]\quad p[5]\}$, and facts b and p protected, i.e. $S = \{b, p\}$. The associated logic program is $P_{PB} = \mathcal{T}(PB) = \{f \leftarrow b, \sim bf$ [1'] $\neg f \leftarrow p, \sim pnf$ [2'] $\perp \leftarrow f, \neg f$ [3] b [4] p [5]$\}$, $least(PB) = \{b, p, f, \neg f, \perp\}$, and $WFM(P_{PB}) = \{f, b, \sim bf, \neg f, p, \sim pnf, \perp\}$.

$CRS_1(P_{PB}) = \{\sim bf\}$ and $CRS_2(P_{PB}) = \{\sim pnf\}$ and the neighbour contradiction free programs are:

$NCFP_1(P_{PB}) = \{f \leftarrow b, \sim bf$ [1'] $\neg f \leftarrow p, \sim pnf$ [2'] $\perp \leftarrow f, \neg f$ [3] $bf \leftarrow \sim bf$ [6] b [4] p [5]$\}$ with $MNS_1(P_{PB}) = WFM_1(P_{PB}) = \{b, p, \sim pnf, \neg f\}$.

$NCFP_2(P_{PB}) = \{f \leftarrow b, \sim bf$ [1'] $\neg f \leftarrow p, \sim pnf$ [2'] $\perp \leftarrow f, \neg f$ [3] $pnf \leftarrow \sim pnf$ [7] b [4] p [5]$\}$ with $MNS_2(P_{PB}) = WFM_2(P_{PB}) = \{b, p, \sim bf, f\}$.

Note that in all maximal non-contradictory submodels MNS_i only b and p hold.

The inverse transformation $\mathcal{T}^{-1}$ may be defined as follows:

Definition 5.6 (Inverse transformation-$\mathcal{T}^{-1}$) Given a logic program P associated with a Program Base $PB = \langle K, S, IC\rangle$ the latter may be obtained as follows:

- For each positive rule $H \leftarrow B_1, \ldots, B_n$ in P there is a similar rule both in K and in S.
- For each non positive rule $H \leftarrow B_1, \ldots, B_n, \sim A$ in P such that $\sim A \in WFM(P)$ there is instead the rule $H \leftarrow B_1, \ldots, B_n$ in K.
- IC is the set of all integrity constraints in P.

Let R denote the set of rules in P other than integrity constraints. Then by the above definition K is the result of the 3-valued GL-transformation [PP90] applied to R and its well founded model (i.e. $K = R/WFM(R)$).

Given a $PB = \langle K; S; IC\rangle$, by the **current world** i we mean the model $least(K \cup IC)$ or $WFM(P_{PB})$ which may be contradictory.

5.1 Minimal Assumption Revision

Definition 5.7 (Counterfactual value) Given a $PB = \langle K, S, IC \rangle$ and $P_{PB} = \mathcal{T}(PB)$ we define a counterfactual implication $\phi > \psi$ to be true, iff ψ holds in all maximal non-contradictory submodels of $P_{PB} \cup \{\phi\}$, i.e.

$$\psi \in \cap_i MNS_i(\mathcal{T}(P_{PB} \cup \{\phi\}))$$

Definition 5.7 implicitly specifies the set of most similar ϕ-worlds as follows:

Definition 5.8 (Most similar ϕ-worlds) Given a Program Base PB with associated logic program P_{PB}, the set of most similar ϕ-worlds $f(\phi, i)$ is defined as:

$$f(\phi, i) = \{M | M = MNS(\mathcal{T}(PB) \cup \{\phi\})\}$$

This means that the most-similar worlds to be considered are defined by the CR Semantics by removing contradiction from the associated logic program.

The selection function $f(\phi, i)$ in the above definition may also be equivalently defined as[PAA91a]:

$$f(\phi, i) = \{M | M = WFM(NCFP(\mathcal{T}(PB) \cup \{\phi\}))\}$$

or, by definition:

$$f(\phi, i) = \{least(\mathcal{T}^{-1}(S)) | S = NCFP(\mathcal{T}(PB) \cup \{\phi\})\})$$

Since $\mathcal{T}(P_i \cup \{\phi\}) = \mathcal{T}(P_{PB}) \cup \{\phi\}$[3]

$$f(\phi, i) = \{least(\mathcal{T}^{-1}(S)) | S = NCFP(P \cup \{\phi\})\}$$

This means that, given a Program Base PB, in order to obtain a counterfactual implication $\phi \rightarrow \psi$, first obtain program P, then add $\{\phi\}$ and evaluate the maximal non-contradictory submodels of the new program. The counterfactual is true iff ψ holds in all such maximal non-contradictory submodels, which are the most similar ϕ-worlds of i.

Remark 5.1 Note that the transformation result does not depend on any particular counterfactual antecedent, i.e. the associated logic program P is independent of the counterfactual being evaluated. This plays an important role if we consider the logic program as an implementation of the Program Base. Moreover, for any Program Base PB_1 and PB_2 we have $\mathcal{T}(PB_1 \cup PB_2) = \mathcal{T}(PB_1) \cup \mathcal{T}(PB_2)$, a modular construction.

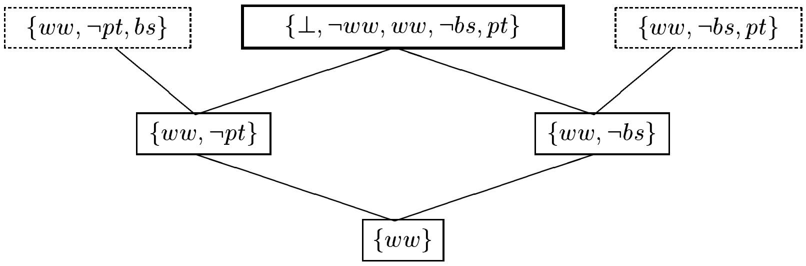
Example 3 Consider a variant of example 2 with $S = \{b, p\}$, $P_{PB} = \{f \leftarrow b[1] \quad \neg f \leftarrow p\ [2] \quad \bot \leftarrow f, \neg f\ [3]\ b\ [4]\}$. The counterfactual $p > \neg f$ is shown to be false since $P_2 = P_1$ and the most similar worlds where p holds are $f(p, i) = \{\{b, p, \neg f\}, \{b, p, f\}\}$. By the same token the counterfactual $b > f$ is false.

Example 4 Consider the counterfactual premise ww and the Program Base $PB = \langle \{\neg ww \leftarrow \neg pt, \neg bs\ [1] \quad \neg pt\ [2] \quad \neg bs\ [3]\}; \{\}; \{\neg ww \leftarrow \neg pt, \neg bs\} \rangle$ with associated logic program $P_{PB} = \{\neg ww \leftarrow \neg pt, \neg bs\ [1] \quad \neg pt \leftarrow \sim pt'\ [2] \quad \neg bs \leftarrow \sim bs'\ [3]\}$, meaning "The wheel of my bycicle is not wobbly, whenever I don't prove I have a punctured tube, nor a broken-spoke".

[3]This follows easily from the definiton of $\mathcal{T}$.

The introduction in P_{PB} of CWA-inhibition rules of the form $L \leftarrow \sim L$ to remove $\perp$ accounts for the struture shown in the figure below (where only the relevant literals are shown), except for the dashed models. When evaluating the counterfactual $ww > bs$ both maximal non-contradictory submodels of the contradictory one are considered, and hence the counterfactual implication is false.

For the evaluation of the existencial counterfactual $ww >_\exists bs$, in definition 4.5 besides rule $L \leftarrow \sim L$ we add also the two rules $\neg L \leftarrow \sim L, L \leftarrow \sim \neg L$. Note that these rules do not change the WFM when added together. It can be shown that the dashed models above the maximal non-contradictory submodels exist. They are most-similar worlds to the contradictory model since in them a minimal number of CWA-assumptions were revised. The counterfactual $ww >_\exists bs$ (resp. $ww >_\exists pt$) is true, since a ww model with bs, (resp. pt) exists. The dashed models are in fact Extended Stable Models (XSMs) of the programs obtained with the revised definition 4.5. The new definition does not alter the previous results for non existencial counterfactuals since the WFM, being the intersection of all XSMs, expresses what holds in all of them. For the purpose of what holds in all most similar worlds it is enough to consider the WFMs obtained according to the original definition 4.5. The MNSs can be seen as partial worlds with respect to the CWA literals, and the two extra rules then allow their completion in that respect.



6 Lewis's similarity precepts obeyed

In this section we prove that the notion of similarity introduced in definition 5.8 satisfies lewis preceipts of similarity.

Proposition 1 Let $PB = \langle K; IC; S \rangle$ be a Program Base and a an atom. Then $a \in least(K \cup IC)$ iff $a \in WFM(\mathcal{T}(PB))$.

Corollary 6.1 A Program Base PB is contradictory iff $\mathcal{T}(PB)$ is contradictory.

Proposition 2 Let P be a logic program, IC a set of integrity constraints and a an atom. Then $a \in WFM(P)$ iff $a \in least(\mathcal{T}^{-1}(P))$.

Corollary 6.2 A program P is contradictory w.r.t. a set of ICs iff $\mathcal{T}^{-1}(P)$ is contradictory.

Proposition 3 Let $PB = \langle K; IC; S \rangle$ be a Program Base and a an atom. Then $\mathcal{T}(PB) \cup \phi = \mathcal{T}(\langle P \cup \{\phi\}; IC; S \cup \{\phi\} \rangle)$.

Proof Follows directly from the definition of $\mathcal{T}$. $\square$

In the sequel we consider the current world (i.e. the $WFM(P_{PB})$) to be non-contradictory, and by a contradiction removal set we always mean a minimal contradiction removal set.

Lewis's selection function properties

i) If ϕ is true at i, then $f(\phi, i)$ is the set $\{i\}$ having i as its only member.

Proof As $i \models \phi$ then $least(i) = least(i \cup \{\phi\})$. So $i \cup \{\phi\}$ is non-contradictory. By proposition 1, the program $\mathcal{T}(i \cup \{\phi\})$ is also non-contradictory. According to the properties of CRS its only $NCFP$ is itself so:

$$f(\phi, i) = \{least(\mathcal{T}^{-1}(\mathcal{T}(i \cup \{\phi\})))\}$$

which by, proposition 2, is equal to $\{i\}$. $\Box$

ii) $f(\phi, i)$ is included in $[\![\phi]\!]$.

Proof According to CRS every fact of a program belongs to its $CRWFM$ (and to all its MNS's). As $\mathcal{T}(\phi \cup \{i\})$ has ϕ as a fact, ϕ belongs to all its MNSs. By proposition 2 elements of $f(\phi, i)$ have ϕ, i.e. are ϕ-worlds. $\Box$

iii) If $[\![\phi]\!]$ is included in $[\![\psi]\!]$ and $f(\phi, i)$ is nonempty, then $f(\psi, i)$ is also nonempty.

Proof (1) - If $\phi \in least(i)$, i.e. $i \in [\![\phi]\!]$: by hypothesis $i \subseteq [\![\psi]\!]$ and by proposition 1 $f(\psi, i) = \{\psi\}\{\}$.
(2) - If $\phi \notin least(i)$: by hypothesis $\exists S_r$ s.t. $\mathcal{T}(i \cup \{\phi\}) \cup S_r$ is a non-contradictory program (where S_r is a set of CWA-inhibition rules). Then:

$\exists S_r$ $\mathcal{T}(i) \cup \{\phi\} \cup S_r$ is non-contradictory.
$\exists S_r$ $\mathcal{T}(i) \cup \{\psi\} \cup \{\phi\} \cup S_r$ is non-contradictory by the first hypothesis.
$\exists S_r$ $\mathcal{T}(i \cup \{\psi\}) \cup \{\phi\} \cup S_r$ is non-contradictory.

For $\mathcal{T}(i \cup \{\psi\}) \cup S_r$ to be contradictory there should exist one $S = SS(\bot)$ s.t. $\sim\phi \in SS(\bot)$ in $\mathcal{T}(i \cup \{\psi\}) \cup S_r$. By definition of CWA-inhibition rule none of the S_r literal sbelong to the WFM. So for $\sim\phi$ to belong to S it must belong to $\mathcal{T}(i \cup \{\psi\})$. By definition of $\mathcal{T}$, the only literals negative in $WFM(\mathcal{T}(PB))$ that appears in a body of any rule in $\mathcal{T}(PB)$ are new atoms. So $\sim\phi$ cannot be in any SS of any positive literal (in particular of $\bot$). Thus $\exists S_r$ s.t. $\mathcal{T}(i \cup \{\psi\}) \cup S_r$ is non-contradictory. $\Box$

iv) If $[\![\phi]\!] \subseteq [\![\psi]\!]$ and $f(\psi, i) \cap [\![\phi]\!]\{\}$ then $f(\phi, i) = f(\psi, i) \cap [\![\phi]\!]$.

Proof If $f(\psi, i) \cap [\![\phi]\!]$ is non-empty then $f(\psi, i)\{\}$ and
$f(\psi, i) \cap [\![\phi]\!] = \{least(\mathcal{T}^{-1}(S)) | S = NCFP(\mathcal{T}(i \cup \{\psi\}))$ and $\phi \in WF(S)\} =$
$= \{least(\mathcal{T}^{-1}(S)) | S = NCFP(\mathcal{T}(i \cup \{\psi\}) \cup \{\phi\})\}$ by $WF(S) = WF(S \cup \{\phi\})$
$= \{least(\mathcal{T}^{-1}(S)) | S = NCFP(\mathcal{T}(i) \cup \{\psi\}) \cup \{\phi\}\}$
$= \{least(\mathcal{T}^{-1}(S)) | S = NCFP(\mathcal{T}(i \cup \{\phi\}))\}$ by $\phi \models \psi$
$= f(\phi, i)$ by definition $\Box$

7 Conclusion

We presented a new approach to counterfactual evaluation based on a notion of assumption revision. We showed through examples that CRS semantics is adequate to formalize this form of common-sense reasoning, thus complementing the work presented in [Prz89, PAA91d, PAA91c]. It remains open how this notion of similarity relates to other proposals such as in [Gin86, Win88, Jac89].

Acknowledgements

We thank ESPRIT BRA COMPULOG (no. 3012), Instituto Nacional de Investigao Cientfica, Junta Nacional de Investigao Cientfica and Gabinete de Filosofia do Conhecimento for their support.

References

[Gin86] M. Ginsberg. Counterfactuals. *Artificial Intelligence*, 30(1):35–79, 1986.

[GL90] M. Gelfond and V. Lifschitz. Logic programs with classical negation. In *ICLP'90*, pages 579–597. MIT Press,1990.

[Jac89] P. Jackson. On the semantics of counterfactuals. In *IJCAI89*, pages 1382–1387, 1989.

[Lew73] D. Lewis. *Counterfactuals.* Harvard, 2nd edition, 1973.

[PAA91a] L. M. Pereira, J. J. Alferes, and J. N. Aparício. Contradiction Removal within well founded semantics. In Anil Nerode, Wiktor Marek, and V. S. Subrahmanian, editors, *1st International Workshop on Logic Programming and Non-monotonic Reasoning (to appear).* MIT Press, 1991.

[PAA91b] L. M. Pereira, J. J. Alferes, and J. N. Aparício. A practical introduction to well founded semantics. In *Third Scandinavian Conference on Artificial Intelligence (to appear).* IOS Press, 1991.

[PAA91c] L. M. Pereira, J. N. Aparício, and J. J. Alferes. Hypothetical reasoning within well founded semantics. In *Third Scandinavian Conference on Artificial Intelligence (to appear).* IOS Press, 1991.

[PAA91d] L. M. Pereira, J. N. Aparício, and J. J. Alferes. Non-monotonic reasoning with well founded semantics. In *International Conference on Logic Programming (to appear).* MIT Press, 1991.

[PP90] H. Przymusinska and T. Przymusinski. *Semantic Issues in Deductive Databases and Logic Programs.* Formal Techniques in Artificial Intelligence. North Holland, 1990.

[Prz89] T. C. Przymusinski. Three-valued formalizations of non-monotonic reasoning and logic programming. In H. Levesque R. Brachman and R. Reiter, editors, *1st Conference on Principles of Knowledge Representation and Reasoning*, pages 341–348. Morgan Kaufmann Publishers, 1989.

[Prz90] T. Przymusinski. Extended stable semantics for normal and disjunctive programs. In *ICLP'90*, pages 459–477. MIT Press, 1990.

[Res64] N. Rescher. *Hypothetical Reasoning.* North-Holland, 1964.

[VEK76] A. Van Emden and R. Kowalski. The semantics of predicate logic as a programming language. *JACM*, 23(4):733–742, 1976.

[VGRS90] A. Van Gelder, K. A. Ross, and J. S. Schlipf. The well-founded semantics for general logic programs. *JACM*, 1990.

[Win88] M. Winslett. Reasoning about action using a possible models approach. In *AAAI88*, pages 89–93. Morgan Kaufmann, 1988.