Alternative Characterizations for Strong Equivalence of Logic Programs

Pedro Cabalar

AI Lab., Dept. of Computer Science
University of Corunna, Spain
cabalar@dc.fi.udc.es

Abstract

In this work we present additional results related to the property of strong equivalence of logic programs. This property asserts that two programs share the same set of stable models, even under the addition of new rules. As shown in a recent work by Lifschitz, Pearce and Valverde, strong equivalence can be simply reduced to equivalence in the logic of *Here-and-There* (HT). In this paper we provide two alternatives respectively based on classical logic and 3-valued logic. The former is applicable to general rules, but not for nested expressions, whereas the latter is applicable for nested expressions but, when moving to an unrestricted syntax, it generally yields different results from HT.

1 Introduction

There is no doubt that the application of logic programming (LP) as a tool for knowledge representation has influenced in the progressive evolution of LP towards a more logical-style orientation, avoiding the initial syntactic restrictions. Think, for instance, how the stable models semantics [4] has been successively modified to cope with explicit negation and disjunctive heads [5], default negation in the head [9, 6] or, finally, the full use of nested expressions [11]. Perhaps as a result of this evolution, the following question has become interesting: when can we consider that two (syntactically different) programs Π_1 and Π_2 *represent the same knowledge*?

From a traditional LP perspective, we would say that Π_1 and Π_2 are equivalent when they share the same set of stable models like, for instance, the programs $\{p\}$ and $\{p \leftarrow \mathit{not}\ q\}$ whose only stable model is $\{p\}$. However, nonmonotonicity may cause them to behave in a different way in the presence of additional rules (just add fact q to both programs). Thus, if we want to check whether Π_1 and Π_2 actually represent the *same knowledge*, we must require a stronger condition, talking instead about *strong equivalence*: for any Π, the stable models of $\Pi_1 \cup \Pi$ and $\Pi_2 \cup \Pi$ coincide.

An elegant solution to this problem is the recent characterization of stable models relying on Heyting's logic of *Here-and-There* (HT). In [14], Pearce first showed that stable models can be simply seen as some kind of minimal HT models. Then, in [10], Lifschitz, Pearce and Valverde proved that, in fact, this characterization fits with the semantics for nested operators independently proposed in [11] and, what is more important, that two programs are strongly equivalent iff they have the same set of HT models.

In this paper we provide two closely related alternatives to HT that rely on classical logic and 3-valued logic (L_3), respectively. These alternatives present the advantage of using very well-known formalisms, which may help for a better insight of strong equivalence (the main emphasis of this paper), but can be useful for implementation purposes too. Unfortunately, we also show how, in both cases, their scope of applicability seems to be smaller than in the HT case. This is evident for the classical encoding we present, which can only be used as a "direct" semantics[1] for non-nested logic programs, whereas the L_3 characterization properly handles nested expressions in a direct way, but loses some important properties when nesting is also allowed for rule conditionals.

The paper is structured as follows. The next section recalls the basic definition of stable models for general (non-nested) logic programs. Section 3 describes the

[1] Application of the classical encoding to nested expressions is also possible, but only after a previous *syntactic* transformation.

classical encoding. In Sections 4 and 5 we respectively describe nested expressions and the 3-valued formalization. After that, we briefly comment the differences between the HT and L_3 interpretations. Finally, Section 7 concludes the paper. Proofs of theorems have been included in an appendix.

2 Stable models

The syntax of logic programs is defined starting from a finite set of ground atoms Σ, called the *Herbrand base*, which will serve as propositional signature. We assume that all the variables have been previously replaced by their possible ground instances. Letters a, b, c, d, p, q will be used to denote atoms in Σ, and letters I, J to denote subsets of Σ. A *logic program* is defined as a collection of rules of the shape:

$$a_1; \ldots; a_m; \mathit{not}\, b_1; \ldots; \mathit{not}\, b_n \leftarrow c_1, \ldots, c_r, \mathit{not}\, d_1, \ldots, \mathit{not}\, d_s \quad (1)$$

We call *head* (resp. *body*) to the left (resp. right) hand side of the arrow in (1). The comma and the semicolon are alternative representations of conjunction $\wedge$ and disjunction $\vee$, respectively. When $m = n = 0$ we usually write $\bot \leftarrow B$ instead of $\leftarrow B$, whereas when $r = s = 0$ we directly write H instead of $H \leftarrow$ or $H \leftarrow \top$.

Sometimes, it will be convenient to think about program rules as classical propositional formulas, where $\leftarrow$ and *not* are respectively understood as material implication and classical negation. In this way, the usual expression $I \models R$ denotes that interpretation I satisfies rule R (seen as a classical formula), whereas $I \models \Pi$ means that I is a model of the program Π (seen as a classical theory).

The *reduct* of a program Π w.r.t. some set of atoms I, written Π^I, is defined as the result of replacing in Π any default literal *not* p by $\top$, if $p \notin I$, or by $\bot$ otherwise.

Definition 1 (Stable model) *A set of atoms $I \subseteq \Sigma$ is a* stable model *of a logic program Π iff I is a minimal model of Π^I.* □

3 Strong equivalence in classical logic

We can think about the definition of stable models as a try-and-error procedure which handles (propositional) interpretations for two different purposes. On the one hand, we start from some arbitrary interpretation I^a, we can call the initial *assumption*, used to compute the reduct Π^{I^a}. On the other hand, in a second step, we deal with minimal models of Π^{I^a} which, in principle, *need not to have any connection* with I^a. Each minimal model I^p can be seen as the set of propositions we can *prove* by deductive closure using the rules in Π^{I^a}. When the proved atoms coincide with the initial assumption, $I^p = I^a$, a stable model is obtained.

In order to capture this behavior, we reify all the atoms $p \in \Sigma$ to become arguments of two unary predicates, $\mathit{assumed}(p)$ and $\mathit{proved}(p)$, that respectively talk about I^a and I^p. Sort variable X will be used for ranging over any propositional symbol in Σ. When considering the models of any reified formula F, we will implictly assume that they actually correspond to $F \wedge \mathrm{UNA}$, where UNA stands for the unique names assumption for sort Σ. This allows us identifying any Herbrand model M of this type of formulas with a pair[2] (I^p, I^a) so that $M[\mathit{assumed}] = I^a$ and $M[\mathit{proved}] = I^p$. Expression $M \models F$ represents again satisfaction of reified formulas – ambiguity with respect to $I \models F$ is cleared by the shape of structures and formulas.

Given this simple framework, we provide two encodings: the first one is a *completely straightforward* translation to capture stable models, whereas the second one is a stronger translation to characterize strong equivalence.

Definition 2 (First translation) *For any logic program rule R like (1), we define the classical formula $\dot{R}$ as the material implication:*

$$\Big(\bigwedge_{i=1}^{r} \mathit{proved}(c_i)\Big) \wedge \Big(\bigwedge_{i=1}^{s} \neg \mathit{assumed}(d_i)\Big) \supset \Big(\bigvee_{i=1}^{n} \mathit{proved}(a_i)\Big) \vee \Big(\bigvee_{i=1}^{m} \neg \mathit{assumed}(b_i)\Big) \quad (2)$$

Given a logic program Π, the formula $\dot{\Pi}$ stands for the conjunction of all the $\dot{R}$, for each rule $R \in \Pi$. □

Intuitively, to obtain the minimal models I^p of Π^{I^a} we can use an ordering relation among pairs $(I^p, I^a) \preceq (J^p, J^a)$ that holds when both $I^a = J^a$ is fixed and $I^p \subseteq J^p$. The corresponding models $\preceq$-minimization have a simple syntactic counterpart[3]: predicate circumscription $\mathrm{CIRC}[\dot{\Pi}; \mathit{proved}]$. Second, after obtaining minimal models, we must further require $I^p = I^a$, that is, we want pairs of shape (I, I) where what we

[2] The superscripts p and a, which stand here for *proved* and *assumed*, respectively correspond to the worlds *here* and *there* in HT or to the sets of *positive* and *non-negative* atoms in L_3.

[3] See Section 2.5 in [8].

assumed results to be exactly what we proved. These pairs of shape (I, I) will be called *total*. Clearly, forcing models to be total corresponds to including of the formula:

$$\forall X. \ \big(proved(X) \equiv assumed(X)\big) \tag{3}$$

The intuitions above are not new. In fact, they were used in Theorem 5.2 in [13] which, adapted[4] to our current presentation, states the following result:

Proposition 1 *Let Σ be a propositional signature. A set of atoms $I \subseteq \Sigma$ is a stable model of a logic program Π iff $M = (I, I)$ satisfies the formula:*

$$\mathrm{CIRC}[\dot{\Pi}; proved] \wedge (3)$$

$\square$

In order to capture strong equivalence of two programs, it seems that we should not only compare the final selected models, but also the set of non-minimal ones involved in the minimization. For instance, it is easy to see that, due to monotonicity of classical logic, the following proposition trivially applies:

Proposition 2 *Let Π_1 and Π_2 be two logic programs such that $\models \dot{\Pi}_1 \equiv \dot{\Pi}_2$. Then Π_1 and Π_2 are strongly equivalent.* $\square$

Unfortunately, the opposite does not necessarily hold: Π_1 and Π_2 can be strongly equivalent while $\dot{\Pi}_1$ and $\dot{\Pi}_2$ have different models. This is because encoding in Definition 2 allows some models which are actually irrelevant for strong equivalence, as we will show next.

Definition 3 (Subtotal model) *For any reified theory T, a model (I^p, I^a) of T is called* subtotal *iff (I^a, I^a) is also model of T.* $\square$

Let $\mathrm{SUBT}(T)$ represent the set of subtotal models of T (note that total models are also included). It is clear that any model $M \notin \mathrm{SUBT}(\dot{\Pi})$ is irrelevant for selecting the total $\preceq$-minimal models, i.e., for obtaining the stable models of Π. The next theorem shows that the coincidence of subtotal models is a necessary condition for strong equivalence. The proof (in the Appendix) constitutes a direct rephrasing of that for the main theorem in [10].

Theorem 1 *Two logic programs Π_1 and Π_2 are strongly equivalent iff* $\mathrm{SUBT}(\dot{\Pi}_1) = \mathrm{SUBT}(\dot{\Pi}_2)$ $\square$

[4] In [13] they used a duplicated signature (atoms p and p') instead of reification and, therefore, they actually applied parallel circumscription. This result seems to have been first presented in [12].

Theorem 1 points out that the $\dot{\Pi}$ encoding is still too weak for a full characterization of strong equivalence. We show next how, using a more restrictive translation (that is, adding more formulas) it is possible to obtain theories for which all their models are subtotal. To understand how to do this, consider the example program $\Pi_0 = \{p \leftarrow q\}$ where:

$$\dot{\Pi}_0 \quad \stackrel{\mathrm{def}}{=} \quad proved(q) \supset proved(p)$$

This formula has 12 models: it restricts the extent of *proved* to 3 cases ($\emptyset$, $\{p\}$ and $\{p, q\}$) leaving free, in each case, the 4 possibilities for *assumed*.

$$\begin{tikzcd}
(\{p,q\},\{p,q\}) & (\{p\},\{p\}) & (\emptyset,\emptyset) \\
(\{p\},\{p,q\}) \arrow[u] & (\emptyset,\{p\}) \arrow[u] & \\
(\emptyset,\{p,q\}) \arrow[u] & &
\end{tikzcd}$$

Figure 1: Subtotal models of $\dot{\Pi}_0$.

Figure 1 shows the 6 subtotal models of $\dot{\Pi}_0$, representing the $\preceq$-ordering relationships among them. Notice how subtotal models always satisfy $I^p \subseteq I^a$, that is, we can require:

$$\forall X. \big(proved(X) \supset assumed(X)\big) \tag{4}$$

Unfortunatley, the addition of this axiom is still not enough to rule out all the irrelevant models. For instance, $\dot{\Pi}_0 \wedge (4)$ has still one non-subtotal model: $(\emptyset, \{q\})$. This model, however, has the particularity that its assumed atoms $I^a = \{q\}$ do not satisfy the original program rule: $I^a \not\models p \leftarrow q$. As it is well-known, any stable model I of a program Π, is also a classical model: $I \models \Pi$. So, instead of starting from any arbitrary initial assumption I^a, we can begin requiring $I^a \models \Pi$. This can be easily incorporated into the encoding as follows. For each logic program rule R like (1), we define $\ddot{R}$ as:

$$\Big(\bigwedge_{i=1}^{r} assumed(c_i)\Big) \wedge \Big(\bigwedge_{i=1}^{s} \neg assumed(d_i)\Big) \supset$$
$$\Big(\bigvee_{i=1}^{n} assumed(a_i)\Big) \vee \Big(\bigvee_{i=1}^{m} \neg assumed(b_i)\Big) \tag{5}$$

Again, $\ddot{\Pi}$ stands for the conjunction of $\ddot{R}$ for all $R \in \Pi$.

Definition 4 (Second translation) *For any logic program Π we define the formula $\Pi^* \stackrel{\mathrm{def}}{=} \dot{\Pi} \wedge \ddot{\Pi} \wedge (4)$.*
$\square$

The proof for the following theorem uses well-known properties of circumscription (see [8]) to show that the additional formulas *do not affect* to the final set of stable models.

Theorem 2 *For any logic program* Π*:*
$\mathrm{CIRC}[\Pi^*; proved] \wedge (3) \equiv \mathrm{CIRC}[\dot{\Pi}; proved] \wedge (3)$ □

But, of course, the real interest of Π^* is that it finally rules out irrelevant models:

Property 1 *Let* Π *be a logic program. Then, any model* $M = (I^p, I^a)$ *of* Π^* *is subtotal.* □

Finally, this property, together with theorem 1, directly implies:

Theorem 3 *Two logic programs* Π_1 *and* Π_2 *are strongly equivalent iff* $\models \Pi_1^* \equiv \Pi_2^*$. □

4 Nested expressions.

The previous section has shown a way of reducing strong equivalence of logic programs into a simple equivalence test in classical logic, providing in this way a (we think) easier alternative to the HT characterization. However, although we gain in simplicity, it must be noticed that we lose in semantic quality: this classical encoding does not provide a general interpretation for program connectives but, instead, is limited to rules of shape (1). The HT characterization is clearly stronger in this sense, since it provides a direct interpretation for *any* possible nesting of program connectives.

In [11] a more general shape for program rules was considered. A *nested expression* is defined as any propositional combination of atoms with 0-ary operators $\bot$, $\top$, unary operator *not* and binary operators ',' and ';'. A logic program is now a set of rules $Head \leftarrow Body$ where $Head$ and $Body$ are nested expressions (notice that the rule conditional $\leftarrow$ is the only operator that cannot be nested). An example of such a rule could be, for instance:

$$a, b \leftarrow not\,(c; not\,d) \tag{6}$$

Stable models for this kind of programs can be easily described by a simple modification in the definition of program reduct. We define now Π^I as the result of replacing in Π every maximal occurrence[5] of $not\,F$ by $\bot$ if $I \models F$ or by $\top$ otherwise. Note that the previous definition of reduct corresponds to the particular case in which F is an atom.

[5]That is, any $not\,F$ that is not in the scope of an outer *not*.

An interesting result derived from this modified semantics (proposition 7 in [11]) is that any program with nested expressions is strongly equivalent to some (non-nested) program, just consisting of rules like (1). To obtain this non-nested program, the following transformations are defined. Let F, G and H represent nested expressions. By $\alpha \Leftrightarrow \beta$ we mean that we replace some regular occurrence of α by β. Then, we handle the following strongly equivalent transformations:

(i) $F, G \Leftrightarrow G, F$ and $F; G \Leftrightarrow G; F$.

(ii) $(F, G), H \Leftrightarrow F, (G, H)$ and $(F; G); H \Leftrightarrow F; (G; H)$.

(iii) $F, (G; H) \Leftrightarrow (F, G); (F, H)$ and $F; (G, H) \Leftrightarrow (F; G), (F; H)$.

(iv) $not\,(F; G) \Leftrightarrow not\,F, not\,G$ and $not\,(F, G) \Leftrightarrow not\,F; not\,G$.

(v) $not\,not\,not\,F \Leftrightarrow not\,F$.

(vi) $F, \top \Leftrightarrow F$ and $F; \top \Leftrightarrow \top$.

(vii) $F, \bot \Leftrightarrow \bot$ and $F; \bot \Leftrightarrow F$.

(viii) $not\,\top \Leftrightarrow \bot$ and $not\,\bot \Leftrightarrow \top$.

(ix) $(F, G \leftarrow H) \Leftrightarrow (F \leftarrow H), (G \leftarrow H)$.

(x) $(F \leftarrow G; H) \Leftrightarrow (F \leftarrow G), (F \leftarrow H)$.

(xi) $(F \leftarrow G, not\,not\,H) \Leftrightarrow (F; not\,H \leftarrow G)$.

(xii) $(F; not\,not\,G \leftarrow H) \Leftrightarrow (F \leftarrow not\,G, H)$.

For instance, rule (6) can be successively transformed as follows:

$a, b \leftarrow not\,c, not\,not\,d.$ By (iv)

$a \leftarrow not\,c, not\,not\,d,$
$b \leftarrow not\,c, not\,not\,d.$ By (ix)

$a; not\,d \leftarrow not\,c,$
$b; not\,d \leftarrow not\,c.$ By (xi)

This treatment of nested expressions exceeds the applicability of our previous classical logic representation. From a practical point of view, such a limitation is not very important, since we can always unfold nested expressions by applying (i)-(xii). Nevertheless, from a theoretical point of view, this clearly points out that the classical encoding fails as a real semantic characterization for LP connectives.

As shown in [10], one of the important features of the HT formalization, apart from the result for strong equivalence, is that it preserves the above interpretation of nested expressions. We show next that a similar behavior can be obtained using standard 3-valued logic (L_3). Surprisingly, L_3 provides the same interpretation for nested expressions, but generally differs once free nesting of rule conditionals is allowed.

5 L_3: Three valued logic.

We will use propositional syntax plus Lukasiewicz's unary operator[6] $\boldsymbol{l}$. Intuitively, a formula $\boldsymbol{l}F$ is never unknown and points out that F is valuated to true. In this way, $\neg \boldsymbol{l}F$ would mean that "F is not true," i.e., it is either false or unknown. If F, G are L_3 formulas and p an atom of the signature Σ then:

$$p,\ \neg F,\ F \vee G,\ \top,\ \bot,\ \boldsymbol{l}F$$

are also L_3 formulas. Propositional derived operators $(\wedge, \supset, \equiv)$ are defined in the usual way.

A three valued interpretation M is a function $M : \Sigma \longrightarrow \{0, 1/2, 1\}$ assigning to each atom $p \in \Sigma$ a truth value $M(p)$ which can be 0 (false), 1/2 (unknown) or 1 (true). We will usually represent M as the pair of sets of atoms (I^p, I^a) respectively containing the *positive* (true) and *consistent* (non-false) atoms where, of course, we require consistence: $I^p \subseteq I^a$. Consequently:

$$M(p) = \begin{cases} 1 & \text{if } p \in I^p \\ 0 & \text{if } p \notin I^a \\ 1/2 & \text{otherwise} \end{cases}$$

Note that we use here the same notation as for the pairs we handled in the reified approach. This is not casual: the negative information of a 3-valued interpretation will be used to represent default negation, whereas the positive information will represent the set of proved atoms.

Definition 5 (L_3 valuation of a formula) *We extend the valuation function M to any formula F, $M(F) \in \{0, 1/2, 1\}$, so that:*

1) $M(\top) = 1$ and $M(\bot) = 0$

2) $M(\neg F) = 1 - M(F)$

3) $M(F \vee G) = max(M(F), M(G))$

4) $M(\boldsymbol{l}F) = \begin{cases} 1 & \text{if } M(F) = 1 \\ 0 & \text{otherwise} \end{cases}$

□

An interpretation M *satisfies* a formula F, written $M \models_3 F$ when $M(F) = 1$. When F is satisfied by *any* interpretation, we call it an *L_3-tautology* and write $\models_3 F$. As usual, an interpretation is a *model* of a theory when it satisfies all its formulas. Maintaining the previous terminology, a 3-valued interpretation M is called *total* iff it has the shape $M = (I, I)$, that is, it contains no unknown atoms. Clearly, when considering total interpretations, the $\boldsymbol{l}$ operator can be simply removed, and L_3 collapses into 2-valued propositional logic.

LP connectives are simply defined among the following derived operators:

$$\begin{aligned} \boldsymbol{m}F &\stackrel{\text{def}}{=} \neg\boldsymbol{l}\neg F \\ \mathit{not}\, F &\stackrel{\text{def}}{=} \neg\boldsymbol{m}F \\ G \leftarrow F &\stackrel{\text{def}}{=} (\boldsymbol{l}F \supset \boldsymbol{l}G) \wedge (\boldsymbol{m}F \supset \boldsymbol{m}G) \\ F \leftrightarrow G &\stackrel{\text{def}}{=} (F \leftarrow G) \wedge (G \leftarrow F) \end{aligned}$$

It is easy to check that the derived semantics for each one of these operators corresponds to:

5) $M(\boldsymbol{m}F) = \begin{cases} 1 & \text{if } M(F) \neq 0 \\ 0 & \text{otherwise} \end{cases}$

6) $M(\mathit{not}\, F) = \begin{cases} 1 & \text{if } M(F) = 0 \\ 0 & \text{otherwise} \end{cases}$

7) $M(G \leftarrow F) = \begin{cases} 1 & \text{if } M(F) \leq M(G) \\ 0 & \text{otherwise} \end{cases}$

8) $M(F \leftrightarrow G) = \begin{cases} 1 & \text{if } M(F) = M(G) \\ 0 & \text{otherwise} \end{cases}$

Operator $\boldsymbol{m}$ acts is the dual of $\boldsymbol{l}$, ($\boldsymbol{m}F$ can be read as "F is consistent") whereas implication $\leftarrow$ is the one proposed by Fitting [3] and Kunen [7]. When we represent some program Π inside L_3, we will consider it as a single formula consisting in the conjunction of all the program rules. Note also that when $\models_3 F \leftrightarrow G$, we can apply uniform substitution in L_3 as we would do in classical propositional logic. For instance, $\models_3 (\mathit{not}\, F) \leftrightarrow (\bot \leftarrow F)$ means that we can replace any occurrence of $(\mathit{not}\, F)$ by $(\bot \leftarrow F)$ and vice versa. Let $\circ$ and $\bullet$ be two meta-operators, any

[6] For instance, see [1], pag. 8, where $\boldsymbol{l}$ is denoted as □.

of them indistinctly representing $\boldsymbol{l}$ or $\boldsymbol{m}$. Then, the following formulas are also L_3 tautologies:

$$\circ(F \wedge G) \;\leftrightarrow\; \circ F \wedge \circ G \tag{7}$$
$$\circ(F \vee G) \;\leftrightarrow\; \circ F \vee \circ G \tag{8}$$
$$\bullet \circ F \;\leftrightarrow\; \circ F \tag{9}$$
$$\bullet \neg \circ F \;\leftrightarrow\; \neg \circ F \tag{10}$$

As $\boldsymbol{m}$ is defined in terms[7] of $\boldsymbol{l}$, this means that we can unfold any L_3 formula until $\boldsymbol{l}$ is exclusively applied to literals. Using these properties, the following lemma can be easily proved:

Lemma 1 *Let R be a pogram rule like (1), and let the pair $M = (I^p, I^a)$ have the common shape of an L_3-interpretation and a classical interpretation for proved/assumed. Then, $M \models_3 R$ iff $M \models \dot{R} \wedge \ddot{R}$.* □

Besides, by inspection on L_3 semantics, we also have that:

Lemma 2 *For any transformation $\alpha \Leftrightarrow \beta$ in* (i)-(xii)*: $\models_3 \alpha \leftrightarrow \beta$.* □

Theorem 4 *Let Π_1 and Π_2 be two logic programs possibly containing nested expressions. Then Π_1 and Π_2 are strongly equivalent iff: $\models_3 \boldsymbol{l}\Pi_1 \equiv \boldsymbol{l}\Pi_2$.* □

Notice that we check $\boldsymbol{l}\Pi_1 \equiv \boldsymbol{l}\Pi_2$ instead of the stronger condition $\Pi_1 \leftrightarrow \Pi_2$. To understand the difference, consider $\Pi_1 = \{a\}$ and $\Pi_2 = \{a \leftarrow \top\}$. The interpretation $M = (\{a\}, \{a\})$ is the only model of both programs and so, $\models_3 \boldsymbol{l}\Pi_1 \equiv \boldsymbol{l}\Pi_2$. However, $\Pi_1 \leftrightarrow \Pi_2$ is not a tautology, since $M' = (\emptyset, \{a\})$ makes $M'(\Pi_1) = 1/2 \neq 1 = M'(\Pi_2)$.

6 Differences with respect to HT

Theorem 4 shows that HT and L_3 coincide in their interpretations of programs with nested expressions. The next natural question is, do the HT and L_3 interpretations coincide for any arbitrary theory? The answer to this question is negative, as we will show with a pair of counterexamples. Of course, due to theorem 4, these counterexamples cannot be just programs with nested expressions, as defined in Section 4. We study, for instance, a nested conditional, and the negation of a conditional.

Consider the theory consisting of the singleton formula $(a \leftarrow b) \leftarrow c$. In HT, this theory is equivalent to $(a \leftarrow b, c)$, which seems to be the most intuitive solution, whereas in L_3 it is actually equivalent to $(a; \mathit{not}\ c \leftarrow b)$. Both equivalences hold in classical propositional logic. However, for computing stable models, their behavior is quite different. For instance, the theory $\{b, (a \leftarrow c), (c \leftarrow a), ((a \leftarrow b) \leftarrow c)\}$ would have a unique stable model $\{b\}$ under the HT interpretation whereas, under L_3, an additional stable model $\{a, b, c\}$ is obtained.

The second example shows the most important problem of the L_3 interpretation: once we allow arbitrary theories, we may obtain non-subtotal models, something that does not happen[8] in HT. Let Π be the theory $\{b, \mathit{not}\ (a \leftarrow b)\}$. Its unique stable model is $\{b\}$ both in HT and L_3. However, while the pair $(\{b\}, \{b\})$ is the unique HT model[9] of Π, in L_3 there exists a second model $(\{b\}, \{a, b\})$ which is not subtotal. In other words, when using L_3 for this general syntax, the set of L_3 models does not fully characterize strong equivalence.

7 Discussion

The study of strong equivalence is probably one of the most active current topics in research in Logic Programming, as it becomes evident by the increasing amount of new results obtained recently (just to cite three examples [16, 15, 2]).

In [15], a classical logic characterization is also provided, which presents several similarities with the approach we present here. The main difference of Pearce et al's method is that it actually relies on a syntactic translation from HT into classical logic. This translation informally consists in a duplication of the atoms in the signature so that an atom p denotes our *proved* whereas an atom p' would denote *assumed*. In this paper, our initial motivation for using classical logic was to improve the presentation and the understanding. In this way, we have directly started from non-nested programs, trying to capture the definition of stable models in a way as direct as possible. As a result, our characterization does not provide an interpretation of nested connectives. In order to deal with them, we would need to apply a previous step, using transformations (i)-(xii). Pearce et al's encoding starts from HT logic, and so, deals with nested expressions (in the same way as HT does). Besides, the transformation presented in [15] has the additional advantage of being linear, while (i)-(xii) are not polynomial in the general case. Despite of these two advantages of

[7] Of course, we could have equally chosen the dual operator $\boldsymbol{m}$ as the basic one.

[8] See for instance Fact 1 in [10].

[9] In fact, the expression $\mathit{not}\ (a \leftarrow b)$ is HT-equivalent to the pair of constraints $(\bot \leftarrow \mathit{not}\, b)$ and $(\bot \leftarrow a)$.

Pearce et al's approach, it must be noticed that none of the two classical encodings can actually be considered a full-semantics for nested logic programs, since *in both cases*, a previous syntactic transformation is required. Therefore, translation to classical logic is very interesting for practical purposes, but is limited from a purely semantic point of view.

Another similarity between our classical encoding with respect to [15] is, not only how to decide strong equivalence, but how to obtain stable models. In our Section 3 we simply used to that purpose the result presented by Lin and Shoham in [13] and then included slight variations that we proved to be sound. In [15], a quantified boolean formula is used instead:

$$\phi' \wedge \neg\exists V((V < V') \wedge \tau_{HT}[\phi]) \tag{11}$$

where V is the set of atoms, ϕ is the original program, ϕ' results from replacing any atom p by p' and finally $\tau_{HT}[\phi]$ is Pearce et al's translation from HT to classical logic. On the other hand, Lin and Shoham's result involving circumscription can be formulated[10] as:

$$(V = V') \wedge \mathcal{C}[\phi] \wedge \neg\exists V((V < V') \wedge \mathcal{C}[\phi]) \tag{12}$$

where $\mathcal{C}[\phi]$ simply replaces each *not* p by $\neg p'$. Notice how, at least structurally, (12) is very similar to (11).

As for the L_3 encoding, it must also be noticed that other logical characterizations have been obtained apart from HT. In [2], for instance, they use instead another logic, KC, and show that this is, in fact, the weakest intermediate logic (between intuitionistic and classical) that allows capturing strong equivalence of logic programs with nested expressions. An interesting open question is how logic KC deals with nested conditionals since, as we have shown, this is the case where HT and L_3 diverge.

The interest of the encoding we have presented is perhaps the use of a multi-valued logic which is commonly known and understood and yields a natural characterization. We believe that the use of L_3 may be especially suitable as a valid alternative for "by hand" studies of strong equivalence, since it allows applying intuitive properties of $\boldsymbol{l}$ and $\boldsymbol{m}$ operators (as an example, look at the proofs in the appendix) in a more or less intuitive way.

Acknowledgements

I want to thank Vladimir Lifschitz for his discussions and comments about a preliminary draft of the L_3 encoding, and to the anonymous referees for drawing my attention to part of the related work cited in this paper. This research is partially supported by the Government of Spain, grant TIC2001-0393.

[10] As described for instance in [8], propositional circumscription is nothing else but a quantified boolean formula.

References

[1] R. Bull and K. Segerberg. Basic modal logic. In D.M. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, volume 2, pages 1–88. D. Reidel Publishing Company, 1984.

[2] A. Hendriks D.H.J. de Jongh. Characterization of strongly equivalent logic programs in intermediate logics, 2001. Unpublished draft. `http://turing.wins.uva.nl/~lhendrik/` .

[3] M. Fitting. A kripke-kleene semantics for logic programs. *Journal of Logic Programming*, 2(4):295–312, 1985.

[4] M. Gelfond and V. Lifschitz. The stable model semantics for logic programming. In R. A. Kowalski and K. A. Bowen, editors, *Logic Programming: Proc. of the Fifth International Conference and Symposium (Volume 2)*, pages 1070–1080. MIT Press, Cambridge, MA, 1988.

[5] M. Gelfond and V. Lifschitz. Classical negation in logic programs and disjunctive databases. *New Generation Computing*, pages 365–385, 1991.

[6] K. Inoue and C. Sakama. Negation as failure in the head. *Journal of Logic Programming*, 35(1):39–78, 1998.

[7] K. Kunen. Negation in logic programming. *Journal of Logic Programming*, 4(4):289–308, 1987.

[8] V. Lifschitz. Circumscription. In C.J. Hogger D.M. Gabbay and J.A. Robinson, editors, *Handbook of Logic in AI and Logic Programming*, volume 3, pages 298–352. Oxford University Press, 1993.

[9] V. Lifschitz. Minimal belief and negation as failure. *Artificial Intelligence*, 70:53–72, 1994.

[10] V. Lifschitz, D. Pearce, and A. Valverde. Strongly equivalent logic programs. *ACM Transactions on Computational Logic*, 2000. (to appear).

[11] V. Lifschitz, L. R. Tang, and H. Turner. Nested expressions in logic programs. *Annals of Mathematics and Artificial Intelligence*, 25:369–389, 1999.

[12] F. Lin. *A Study of Nonmonotonic Reasoning*. PhD thesis, Stanford, 1991.

[13] F. Lin and Y. Shoham. A logic of knowledge and justified assumptions. *Artificial Intelligence*, 57:271–289, 1992.

[14] D. Pearce. A new logical characterisation of stable models and answer sets. In *Non monotonic extensions of logic programming. Proc. NMELP'96. (LNAI 1216)*. Springer-Verlag, 1997.

[15] D. Pearce, H. Tompits, and S. Woltran. Encodings of equilibrium logic and logic programs with nested expressions. In P. Bradzil and A. Jorge, editors, *Lecture Notes in Artificial Intelligence*, volume 2258, pages 306–320. Springer Verlag, 2001.

[16] H. Turner. Strong equivalence for logic programs and default theories (made easy). In *In Proc. of the Sixth Int'l Conf. on Logic Programming and Nonmonotonic Reasoning (LPNMR'01)*, pages 81–92, 2001.

Appendix A. Proofs.

Proof of theorem 1
The right to left direction was proved in proposition 2. For the left to right direction, assume there is some $M = (I^p, I^a)$, $M \in \mathrm{SUBT}(\dot{\Pi}_1)$ but $M \notin \mathrm{SUBT}(\dot{\Pi}_2)$. Let M' be the total pair $M' = (I^a, I^a)$. We will show that it is possible to construct a program Π so that I^a is a stable model of $\Pi_1 \cup \Pi$ but not of $\Pi_2 \cup \Pi$, or vice versa. We consider two cases:

1. $M' \notin \mathrm{SUBT}(\dot{\Pi}_2)$. As $M \in \mathrm{SUBT}(\dot{\Pi}_1)$, by definition of SUBT, M' is model of $\dot{\Pi}_1$. We construct the program $\Pi = \{p : p \in I^a\}$. Clearly, M' is still model of $\dot{\Pi}_1 \wedge \dot{\Pi}$ but, now, it is also minimal: a strict subset of *proved* atoms is not allowed. Therefore, I^a is a stable model of $\Pi_1 \cup \Pi$. On the other hand, as M' is total, but $M' \notin \mathrm{SUBT}(\dot{\Pi}_2)$, then $M' \not\models \dot{\Pi}_2$. As a result, I^a cannot be stable model of $\Pi_2 \cup \Pi$.

2. $M' \in \mathrm{SUBT}(\dot{\Pi}_2)$. Note first that, as $M \notin \mathrm{SUBT}(\dot{\Pi}_2)$ and $M \preceq M'$ we conclude $M \not\models \dot{\Pi}_2$ and $M \prec M'$. We construct the program:

$$\begin{aligned} \Pi \quad = \quad & \{p : p \in I^p\} \cup \\ & \{(p \leftarrow q) : p, q \in (I^a - I^p), p \neq q\} \end{aligned}$$

   It is easy to see that $M' \models \dot{\Pi}$ and so, $M' \models \dot{\Pi}_2 \wedge \dot{\Pi}$. We will show that M' is also minimal for that theory. Let us assume that there exists another model (J, I^a) strictly lower than M', $J \subset I^a$, and try to reach a contradiction. As atoms of I^p are facts of Π, but $M = (I^p, I^a)$ is not model of $\dot{\Pi}_2$, we must have: $I^p \subset J$. Now, take some $q \in J - I^p$ and some $p \in I^a - J$. The rule $p \leftarrow q$ belongs to Π but (J, I^a) does not satisfy the rule translation. Therefore, $(J, I^a) \not\models \dot{\Pi}$ and we reach a contradiction. Since we have proved that M' is minimal, I^a is stable model of $\dot{\Pi}_2 \cup \dot{\Pi}$. Finally, we study program Π_1. It can be easily checked that $M \models \dot{\Pi}$ and so $M \models \dot{\Pi}_1 \cup \dot{\Pi}$. But as $M \prec M'$, M' is not minimal for $\dot{\Pi}_1 \cup \dot{\Pi}$ and so I^a is not stable model for this program.

Proof of theorem 2
We start from $\mathrm{CIRC}[\Pi^*; \mathit{proved}] \wedge (3)$, observing that the conjuncts $\ddot{\Pi}$ and (4) of Π^* can be moved outside the circumscription. The reason for this is that in $\ddot{\Pi}$, predicate *proved* does not occur, whereas in (4) it occurs negatively (considering $\alpha \supset \beta$ as $\neg\alpha \vee \beta$). Therefore, we can respectively apply equivalence (3.2) and proposition 3.3.1 in [8], to obtain the equivalent formula:

$$\mathrm{CIRC}[\dot{\Pi}; \mathit{proved}] \wedge \ddot{\Pi} \wedge (4) \wedge (3)$$

Now, notice that, by definition, the circumscription of $\mathrm{CIRC}[\dot{\Pi}; \mathit{proved}]$ has the shape of a conjunction $\dot{\Pi} \wedge \alpha$ (where α is a second order formula we do not need to detail). So, the above formula is actually:

$$\dot{\Pi} \wedge \alpha \wedge \ddot{\Pi} \wedge (4) \wedge (3)$$

Finally, notice that (3) implies (4), whereas $\dot{\Pi} \wedge (3)$ implies $\ddot{\Pi}$, and so, the implied conjuncts can be removed, obtaining the equivalent formula:

$$\dot{\Pi} \wedge \alpha \wedge (3)$$

which is nothing else but $\mathrm{CIRC}[\dot{\Pi}; \mathit{proved}] \wedge (3)$.

Proof of property 1
We must show that $M' = (I^a, I^a)$ is also model of Π^*. Clearly, as $\ddot{\Pi}$ exclusively refers to predicate *assumed*, $M \models \ddot{\Pi}$ iff $M' \models \ddot{\Pi}$. On the other hand, it is clear that $M' \models (4)$ – in fact, it further satisfies (3). So, we only have to prove that $M' \models \dot{\Pi}$. Assume that, for some rule $R \in \Pi$, $M' \not\models \dot{R}$, which in classical logic is equivalent to $M' \models \neg(\dot{R})$, i.e., M' must satisfy:

$$\Big(\bigwedge_{i=1}^{r} \mathit{proved}(c_i)\Big) \wedge \Big(\bigwedge_{i=1}^{s} \neg \mathit{assumed}(d_i)\Big) \wedge \\ \Big(\bigwedge_{i=1}^{n} \neg \mathit{proved}(a_i)\Big) \wedge \Big(\bigwedge_{i=1}^{m} \mathit{assumed}(b_i)\Big)$$

As in M', *proved* and *assumed* are equivalent, this implies that M' must also satisfy:

$$(\bigwedge_{i=1}^{r} assumed(c_i)) \wedge (\bigwedge_{i=1}^{s} \neg assumed(d_i)) \wedge$$
$$(\bigwedge_{i=1}^{n} \neg assumed(a_i)) \wedge (\bigwedge_{i=1}^{m} assumed(b_i))$$

But this formula is exactly $\neg(\ddot{R})$ and so, M' cannot be model of $\ddot{\Pi}$ nor Π^*, reaching a contradiction.

Proof of lemma 1

Using properties (7)-(10) and the definitions of derived operators, a general logic program rule like (1) is equivalent to the conjunction of the pair of formulas:

$$\begin{aligned} &\boldsymbol{l}c_1 \wedge \cdots \wedge \boldsymbol{l}c_r \wedge \neg \boldsymbol{m}d_1 \wedge \cdots \wedge \neg \boldsymbol{m}d_s \supset \\ &\boldsymbol{l}a_1 \vee \cdots \vee \boldsymbol{l}a_n \vee \cdots \vee \neg \boldsymbol{m}b_1 \vee \cdots \vee \neg \boldsymbol{m}b_m \end{aligned} \quad \text{(A-1)}$$

$$\begin{aligned} &\boldsymbol{m}c_1 \wedge \cdots \wedge \boldsymbol{m}c_r \wedge \neg \boldsymbol{m}d_1 \wedge \cdots \wedge \neg \boldsymbol{m}d_s \supset \\ &\boldsymbol{m}a_1 \vee \cdots \vee \boldsymbol{m}a_n \vee \cdots \vee \neg \boldsymbol{m}b_1 \vee \cdots \vee \neg \boldsymbol{m}b_m \end{aligned} \quad \text{(A-2)}$$

Note that in formulas, operators $\boldsymbol{l}$ and $\boldsymbol{m}$ are only applied to atoms and, vice versa, atoms only appear in the scope of $\boldsymbol{l}$ and $\boldsymbol{m}$. As the interpretation of $\boldsymbol{l}$ and $\boldsymbol{m}$ is always bivalued, we can just consider the meaning of each $\boldsymbol{l}p$ and $\boldsymbol{m}p$ individually and then interprete the whole formula as a classical propositional one. Now, note that $\boldsymbol{l}p$ is true iff $p \notin I^p$ that is $M \models proved(a)$, and $\boldsymbol{m}p$ is true iff $p \in I^a$, that is, $M \models assumed(a)$. Therefore, (A-1) and (A-2) respectively correspond to $\dot{R}$ and $\ddot{R}$.

Proof of lemma 2

As ',' and ';' are simply understood as $\wedge$ and $\vee$, properties (i)-(iii), (vi) and (vii) are straightforward.

(iv) The formula $not\,(F;G)$ is, by definition, $\neg \boldsymbol{m}(F \vee G)$ which, by (8), is always valuated in the same way as $\neg(\boldsymbol{m}F \vee \boldsymbol{m}G)$. As L_3 satisfies the De Morgan's laws, this is equivalent to $\neg \boldsymbol{m}F \wedge \neg \boldsymbol{m}G$, which by definition, is the formula $not\ F, not\ G$. The other part of property (iv) is analogous.

(v) The formula $not\ not\ not\ F$ is, by definition, $\neg \boldsymbol{m} \neg \boldsymbol{m} \neg \boldsymbol{m} F$. By (10), we replace $\boldsymbol{m} \neg \boldsymbol{m} F$ by $\neg \boldsymbol{m} F$ obtaining $\neg \boldsymbol{m} \neg \neg \boldsymbol{m} F$, that is, $\neg \boldsymbol{m}\boldsymbol{m} F$. Now, by (9) this is equivalent to $\neg \boldsymbol{m} F$, which is, by definition $not\,F$.

(viii) The formula $not\ \bot$ is $\neg \boldsymbol{m}\bot$. As $\boldsymbol{m}\bot$ is always valuated as 0 (false), this formula is equivalent to $\neg\bot$, that is, $\top$. The other part of (viii) is analogous.

(ix) The formula $(F, G \leftarrow H)$ is, by definition:

$$(\boldsymbol{l}H \supset \boldsymbol{l}(F \wedge G)) \wedge (\boldsymbol{m}H \supset \boldsymbol{m}(F \wedge G))$$

Applying (7) we obtain:

$$(\boldsymbol{l}H \supset \boldsymbol{l}F \wedge \boldsymbol{l}G) \wedge (\boldsymbol{m}H \supset \boldsymbol{m}F \wedge \boldsymbol{m}G)$$

This is equivalent to:

$$\begin{aligned} &(\neg \boldsymbol{l}H \vee \boldsymbol{l}F) \wedge (\neg \boldsymbol{l}H \vee \boldsymbol{l}G) \wedge \\ &\qquad (\neg \boldsymbol{m}H \vee \boldsymbol{m}F) \wedge (\neg \boldsymbol{m}H \vee \boldsymbol{m}G) \end{aligned}$$

(simply unfold the definition of $\supset$ and apply distributivity). Finally, note that, by definition of $\leftarrow$, the last formula is just:

$$(F \leftarrow H) \wedge (G \leftarrow H)$$

(ix) We just provide a proof sketch:

$$\begin{array}{ll}
(\boldsymbol{l}(G \vee H) \supset \boldsymbol{l}F) \wedge & \\
\quad (\boldsymbol{m}(G \vee H) \supset \boldsymbol{m}F) & \text{Def. of} \leftarrow \\
 & \\
(\boldsymbol{l}G \vee \boldsymbol{l}H \supset \boldsymbol{l}F) \wedge & \\
\quad (\boldsymbol{m}G \vee \boldsymbol{m}H \supset \boldsymbol{m}F) & (8) \\
 & \\
(\boldsymbol{l}G \supset \boldsymbol{l}F) \wedge (\boldsymbol{m}G \supset \boldsymbol{m}F) \wedge & \\
\quad (\boldsymbol{l}H \supset \boldsymbol{l}F) \wedge (\boldsymbol{m}H \supset \boldsymbol{m}F) & \text{Def. of} \supset \\
 & \\
(F \leftarrow G) \wedge (F \leftarrow H) & \text{Def. of} \leftarrow
\end{array}$$

(x) Again, as a proof sketch:

$$\begin{array}{ll}
(\boldsymbol{l}(G \wedge \neg \boldsymbol{m} \neg \boldsymbol{m}H) \supset \boldsymbol{l}F) \wedge & \\
\quad (\boldsymbol{m}(G \wedge \neg \boldsymbol{m} \neg \boldsymbol{m}H) \supset \boldsymbol{m}F) & \text{Def. of} \leftarrow, not \\
(\boldsymbol{l}(G \wedge \neg\neg \boldsymbol{m}H) \supset \boldsymbol{l}F) \wedge & \\
\quad (\boldsymbol{m}(G \wedge \neg\neg \boldsymbol{m}H) \supset \boldsymbol{m}F) & (10) \\
(\boldsymbol{l}G \wedge \boldsymbol{l}\boldsymbol{m}H \supset \boldsymbol{l}F) \wedge & \\
\quad (\boldsymbol{m}G \wedge \boldsymbol{m}\boldsymbol{m}H \supset \boldsymbol{m}F) & (7) \\
(\boldsymbol{l}G \wedge \boldsymbol{m}H \supset \boldsymbol{l}F) \wedge & \\
\quad (\boldsymbol{m}G \wedge \boldsymbol{m}H \supset \boldsymbol{m}F) & (9) \\
(\boldsymbol{l}G \supset \boldsymbol{l}F \vee \neg \boldsymbol{m}H) \wedge & \\
\quad (\boldsymbol{m}G \supset \boldsymbol{m}F \vee \neg \boldsymbol{m}H) & \text{Def.} \supset \\
F; not\,H \leftarrow G & \text{Def. of} \leftarrow, not
\end{array}$$

(xi) It is analogous to proof for (x).

Proof of theorem 4

First, from lemma 2 the corresponding non-nested program has the same set of L_3 models. Second, lemma 1 shows that this program has the same models than our classical encoding. Therefore, we can directly apply theorem 3.