Relativised Equivalence in Equilibrium Logic and its Applications to Prediction and Explanation: Preliminary Report*

David Pearce[1], Hans Tompits[2], and Stefan Woltran[2]

[1] Universidad Rey Juan Carlos, Computing Science and Artificial Intelligence,
Calle Tulipan s/n, E-28933 Madrid, Spain
Email: davidandrew.pearce@urjc.es

[2] Technische Universität Wien, Institut für Informationssysteme 184/3,
Favoritenstrasse 9-11, A-1040 Vienna, Austria
Email: {tompits,stefan}@kr.tuwien.ac.at

Abstract. For a given semantics, two nonmonotonic theories Π_1 and Π_2 can be said to be equivalent if they have the same intended models and strongly (resp., uniformly) equivalent if for any Σ, $\Pi_1 \cup \Sigma$ and $\Pi_2 \cup \Sigma$ are equivalent, where Σ is a set of sentences (resp., literals). In the general case, no restrictions are placed on the language (signature) of Σ. Relativised notions of strong and uniform equivalence are obtained by requiring that Σ belongs to a specified sublanguage $\mathcal{L}$ of the theories Π_1 and Π_2. For normal and disjunctive logic programs under stable-model semantics, relativised strong and uniform equivalence have been defined and characterised in previous work by Woltran. Here, we extend these concepts to nonmonotonic theories in equilibrium logic and discuss applications in the context of prediction and explanation.

1 Introduction

Equilibrium logic [12] is a general purpose formalism for nonmonotonic reasoning extending the stable-model and answer-set semantics for all the usual classes of logic programs, adhering to the general *answer-set programming* (ASP) paradigm. It is a form of minimal-model reasoning in the non-classical logic of *here-and-there*, which is basically intuitionistic logic restricted to two worlds, "here" and "there", and subsumes all important syntactic extensions considered in ASP, including the addition of strong negation, rules with negation-by-default in their heads, and nested programs, as well as those constructs like cardinality and weight constraints and aggregates that have equivalent representations in the more general syntax of equilibrium logic [4, 5].

Recent research in ASP focuses on advanced notions of program equivalence relevant for program optimisation and modular programming [11, 1, 14]. A traditional concept of equivalence, where two nonmonotonic theories, under a given semantics, are viewed as being equivalent if they have the same intended models, is not adequate for

* This work was partially supported by the Spanish Ministry of Education and Science (MEC) under projects TIC-2003-9001-C02 and TIN2006-15455-CO3, and by the Austrian Science Fund (FWF) under grant P18019.

these purposes because such a notion does not satisfy a replacement property like in classical logic. Better candidates, however, are strong and uniform equivalence. While the former meets a replacement principle by definition, the latter is suitable for hierarchically ordered modules. In formal terms, two nonmonotonic theories, Π_1 and Π_2, are strongly (resp., uniformly) equivalent if for any Σ, $\Pi_1 \cup \Sigma$ and $\Pi_2 \cup \Sigma$ are equivalent, where Σ is a set of sentences (resp., literals). In the general case, no restrictions are placed on the language (signature) of Σ. *Relativised notions* of strong and uniform equivalence are obtained by requiring that Σ belongs to a specified sublanguage $\mathcal{L}$ of the theories Π_1 and Π_2. For normal and disjunctive logic programs under stable-model semantics, relativised strong and uniform equivalence have been defined and characterised in previous work by Woltran [19], together with a discussion about complexity issues and implementation strategies. Furthermore, relativised strong and uniform equivalence are special cases of *update equivalence* introduced by Inoue and Sakama [7].

In this paper, we extend the work of Woltran [19] and Pearce and Valverde [14] by characterising relative notions of equivalence for arbitrary (propositional) theories in equilibrium logic. Furthermore, we discuss how relativised equivalences can be applied to certain problems from the areas of diagnosis and abduction, with respect to the problem of deciding whether two logical descriptions have the same explanatory power, and provide a semantical characterisation of this problem. The formal model of an abductive explanation our discussion is based is an extension of a corresponding concept used by Inoue and Sakama [8] for disjunctive logic programs with default negation in their heads. Finally, we address the computational complexity of relative equivalence in equilibrium logic, showing that it remains on the same level as for logic programs.

2 Equilibrium Logic

We work in the nonclassical logic of here-and-there with strong negation $\mathbf{N}_5$ and its nonmonotonic extension, equilibrium logic [12], which generalises the answer-set semantics for logic programs to arbitrary propositional theories [11]. For more details, the reader is referred to [12, 13] and the logic texts cited below.

Formulas of $\mathbf{N}_5$ are built-up in the usual way using the logical constants $\wedge$, $\vee$, $\rightarrow$, $\neg$, $\sim$, standing respectively for conjunction, disjunction, implication, weak (or intuitionistic) negation, and strong negation. The axioms and rules of inference for $\mathbf{N}_5$ include those of intuitionistic logic (see, e.g., [16]) and the strong negation axioms from the calculus of Vorob'ev [17, 18]; for details, see [13].

The model theory of $\mathbf{N}_5$ is based on the usual Kripke semantics for Nelson's constructive logic $\mathbf{N}$ (see, e.g., [6, 16]), but $\mathbf{N}_5$ is complete for Kripke frames $\mathcal{F} = \langle W, \leq \rangle$ (where as usual W is the set of *points* or *worlds* and $\leq$ is a partial-ordering on W) having exactly two worlds, say h ("here") and t ("there") with $h \leq t$. As usual, a *model* is a frame together with an assignment i that associates to each element of W a set of *literals*[1] such that if $w \leq w'$ then $i(w) \subseteq i(w')$. An assignment is then extended inductively to all formulas via the usual rules for conjunction, disjunction, implication and (weak) negation in intuitionistic logic together with the following rules governing

[1] We use the term "literal" to denote an atom, or an atom prefixed by strong negation.

strongly negated formulas:

$$\begin{array}{l}\sim(\varphi \wedge \psi) \in i(w) \text{ iff } \sim\varphi \in i(w) \text{ or } \sim\psi \in i(w);\\ \sim(\varphi \vee \psi) \in i(w) \text{ iff } \sim\varphi \in i(w) \text{ and } \sim\psi \in i(w);\\ \sim(\varphi \to \psi) \in i(w) \text{ iff } \varphi \in i(w); \text{ and } \sim\psi \in i(w);\\ \sim\neg\varphi \in i(w) \text{ iff } \sim\sim\varphi \in i(w) \text{ iff } \varphi \in i(w).\end{array}$$

It is convenient to represent an $\mathbf{N}_5$-model as an ordered pair $\langle H, T\rangle$ of sets of literals, where $H = i(h)$ and $T = i(t)$ under a suitable assignment i. By $h \leq t$ it follows that $H \subseteq T$. Again, by extending i inductively we know what it means for an arbitrary formula φ to be true in a model $\mathcal{M} = \langle H, T\rangle$. We write $\mathcal{M}, w \models \varphi$ to express that φ is true at world w in model $\mathcal{M}$.

A formula φ is true in a here-and-there model $\mathcal{M} = \langle H, T\rangle$, in symbols $\mathcal{M} \models \varphi$, if it is true at each world in $\mathcal{M}$. A formula φ is said to be *valid* in $\mathbf{N}_5$, in symbols $\models \varphi$, if it is true in all here-and-there models. Logical consequence for $\mathbf{N}_5$ is understood as follows: φ is said to be an $\mathbf{N}_5$-*consequence* of a set Π of formulas, written $\Pi \models \varphi$, iff for all models $\mathcal{M}$ and any world $w \in \mathcal{M}$, $\mathcal{M}, w \models \Pi$ implies $\mathcal{M}, w \models \varphi$. Equivalently, this can be expressed by saying that φ is true in all models of Π. Further properties of $\mathbf{N}_5$ are studied in [10].

Equilibrium models are special kinds of minimal $\mathbf{N}_5$ Kripke models. We first define a partial ordering $\trianglelefteq$ on $\mathbf{N}_5$ models that will be used both to characterise the equilibrium property as well as the property of uniform equivalence.

Definition 1. *Given any two models $\langle H, T\rangle$, $\langle H', T'\rangle$, we set $\langle H, T\rangle \trianglelefteq \langle H', T'\rangle$ if $T = T'$ and $H \subseteq H'$.*

Definition 2. *Let Π be a set of $\mathbf{N}_5$ formulas and $\langle H, T\rangle$ a model of Π.*

1. *$\langle H, T\rangle$ is said to be* total *if $H = T$ (otherwise, if $H \subset T$, it is* non-total*).*
2. *$\langle H, T\rangle$ is said to be an* equilibrium *model if it is total and minimal under $\trianglelefteq$ among models of Π.*

In other words, a model $\langle H, T\rangle$ of Π is in equilibrium if it is total and there is no model $\langle H', T\rangle$ of Π with $H' \subset H$. Equilibrium logic is the logic determined by the equilibrium models of a theory. It generalises answer-set semantics in the following sense: For all the usual classes of logic programs, including normal, extended, disjunctive and nested programs, equilibrium models correspond to answer sets [12, 11]. The "translation" from the syntax of programs to $\mathbf{N}_5$ propositional formulas is the trivial one, viz., a ground rule of an (extended) disjunctive program of the form

$$K_1 \vee \ldots \vee K_k \leftarrow L_1, \ldots L_m, notL_{m+1}, \ldots, notL_n,$$

where the L_i and K_j are literals, corresponds to the $\mathbf{N}_5$ sentence

$$L_1 \wedge \ldots \wedge L_m \wedge \neg L_{m+1} \wedge \ldots \wedge \neg L_n \to K_1 \vee \ldots \vee K_k.$$

A set of $\mathbf{N}_5$ sentences is called a *theory*. Two theories are *equivalent* if they have the same equilibrium models.

3 Relativised Equivalence Concepts

We consider theories Π_1, Π_2, etc., and languages $\mathcal{L}$, $\mathcal{L}'$, etc. It will be convenient notationally viewing a language as a set of literals. A theory is said to be *in* the language $\mathcal{L}$ if all its atomic formulas belong to $\mathcal{L}$.

Definition 3. *Let Π_1 and Π_2 be theories.*

(*i*) *Π_1 and Π_2 are* strongly equivalent relative to $\mathcal{L}$ *iff for any* (*empty or non-empty*) *set Σ of $\mathcal{L}$ formulas, $\Pi_1 \cup \Sigma$ and $\Pi_2 \cup \Sigma$ are equivalent, i.e., have the same equilibrium models.*
(*ii*) *Π_1 and Π_2 are* uniformly equivalent relative to $\mathcal{L}$ *iff for any* (*empty or non-empty*) *set X of $\mathcal{L}$ literals, $\Pi_1 \cup X$ and $\Pi_2 \cup X$ are equivalent, i.e., have the same equilibrium models.*

Note that if the theories are logic programs, this means they have the same answer sets.

We explain some terminology and notation. A model $\langle H, T\rangle$ of a theory Π is said to be *maximally non-total* (or just *maximal*) if it is non-total and is maximal among models of Π under the ordering $\trianglelefteq$. In other words, a model $\langle H, T\rangle$ of Π is maximal if for any model $\langle H', T\rangle$ of Π, if $H \subset H'$ then $H' = T$. It is clear that if a theory Π is finite and has a non-total model $\langle H, T\rangle$, then it has a maximally non-total model $\langle H', T\rangle$ such that $H \subseteq H'$. However, maximal models need not exist in case that Π is an infinite theory. In what follows, we shall assume that all theories are finite.

Let $\mathcal{L}$ be a sublanguage of $\mathcal{L}'$. If $\mathcal{M} = \langle H, T\rangle$ is an $\mathcal{L}'$ model, its *$\mathcal{L}$-1-reduct* is defined by

$$\langle H \cap \mathcal{L}, T\rangle$$

and denoted by $\mathcal{M}|\mathcal{L}$. The term “1-reduct” stems from the fact that it refers to the first component of the model.

4 Characterising Relative Equivalence

For logic programs, the above relativised notions of equivalence are characterised by Woltran [19] in terms of what are called *relativised strong* (resp., *uniform*) *equivalence models*, or *RSE* (resp., *RUE*) *models* for short. We start by re-expressing these concepts in terms of ordinary models in the logic $\mathbf{N}_5$.

Definition 4. *Let Π be a theory in $\mathcal{L}'$ and $\mathcal{L}$ a sublanguage of $\mathcal{L}'$. A model $\mathcal{M} = \langle H, T\rangle$ is an* RSE$_{\mathcal{L}}$-model *of Π if it meets the following criteria:*

4.1 $\mathcal{M}$ is a total model of Π or
4.2 $\mathcal{M}$ is the $\mathcal{L}$-1-reduct of a non-total model $\langle H', T\rangle$ of Π, and
4.3 for any non-total model $\langle J, T\rangle$ of Π, $T \setminus J \cap \mathcal{L} \neq \emptyset$.

In other words, 4.3 holds together with one of 4.1 or 4.2. It is easy to see that for disjunctive logic programs, the above concept coincides with that of an RSE-model as defined by Woltran [19]. Indeed, we must check a preliminary condition and Conditions (i)-(iii) of Definition 6 by Woltran [19]. Clearly, both 4.1 and 4.2 above imply that T is a classical model of Π as required by (i). Condition 4.3 above re-expresses Clause (ii), while Condition 4.2 re-expresses Clause (iii). Finally, we check the preliminary condition of Woltran [19]. By 4.2, if $H \neq T$ then $\langle H, T \rangle$ is the reduct of a non-total model $\langle H', T \rangle$ of Π, so $H' \subset T$. Therefore, $H' \cap \mathcal{L} \subseteq T \cap \mathcal{L}$. But by 4.3, $H' \cap \mathcal{L} \neq T \cap \mathcal{L}$. Since $H = H' \cap \mathcal{L}$, it follows that $H \subset T \cap \mathcal{L}$ as required by the original definition of an RSE-model.

Now the following lemma is straightforward but useful. It says that two models with the same $\mathcal{L}$-1-reduct satisfy the same $\mathcal{L}$-sentences.

Lemma 1. *Let $\mathcal{M}$ and $\mathcal{M}'$ be $\mathbf{N}_5$ models and φ a formula all of whose atoms belong to the language $\mathcal{L}$. If $\mathcal{M}|\mathcal{L} = \mathcal{M}'|\mathcal{L}$, then $\mathcal{M} \models \varphi$ iff $\mathcal{M}' \models \varphi$.*

4.1 Relativised Strong Equivalence

Relativised strong equivalence (RSE) is defined as Woltran [19] does but for arbitrary theories. We can now show that sameness of RSE-models is a sufficient condition to ensure RSE.

Theorem 1. *Let Π_1 and Π_2 be theories having the same RSE$_\mathcal{L}$-models. Then, Π_1 and Π_2 are strongly equivalent relative to $\mathcal{L}$.*

Proof. Assume the hypothesis of the theorem and consider the theory $\Pi_1 \cup \Sigma$ where Σ is any set of sentences in $\mathcal{L}$. Consider any equilibrium model $\mathcal{M} = \langle T, T \rangle$ of $\Pi_1 \cup \Sigma$. We shall show that $\mathcal{M}$ is also an equilibrium model of $\Pi_2 \cup \Sigma$. By the symmetry of the situation, the same argument will show that any equilibrium model of $\Pi_2 \cup \Sigma$ must be an equilibrium model of $\Pi_1 \cup \Sigma$.

We first show that $\mathcal{M}$ is an RSE$_\mathcal{L}$-model of Π_1. Evidently, it is a total model of Π_1, so Condition 4.1 holds. Suppose that Condition 4.3 fails, so that there is a model $\langle J, T \rangle$ of Π_1 with $J \subset T$ such that $T \cap \mathcal{L} = J \cap \mathcal{L}$. Since $\langle T, T \rangle \models \Sigma$, by Lemma 1, $\langle J, T \rangle \models \Sigma$, but this contradicts the assumption that $\langle T, T \rangle$ is an equilibrium model of $\Pi_1 \cup \Sigma$. So Condition 4.3 applies and $\mathcal{M}$ is an RSE$_\mathcal{L}$-model of Π_1 and hence by assumption of Π_2. Therefore

$$\langle T, T \rangle \models \Pi_2 \cup \Sigma.$$

We need to show that it is in equilibrium. Note that since $\langle T, T \rangle$ is an RSE$_\mathcal{L}$-model of Π_2, by Condition 4.3 there is no model $\langle J, T \rangle$ of Π_2 with $J \subset T$ such that $T \cap \mathcal{L} = J \cap \mathcal{L}$. Suppose that $\mathcal{M}$ is not an equilibrium model of $\Pi_2 \cup \Sigma$. Then $\Pi_2 \cup \Sigma$ has a model $\langle H, T \rangle$ with $H \subset T$, so in particular $\langle H, T \rangle \models \Pi_2$ and by 4.3, $T \backslash H \cap \mathcal{L} \neq \emptyset$. So, $H \cap \mathcal{L} \subset T \cap \mathcal{L} \subseteq T$. It follows that $\langle H \cap \mathcal{L}, T \rangle$ is the $\mathcal{L}$-1-reduct of a model $\langle H, T \rangle \models \Pi_2$, with $H \subset T$. By Condition 4.2, $\langle H \cap \mathcal{L}, T \rangle$ is therefore an RSE$_\mathcal{L}$-model of Π_2, hence of Π_1. So, again by 4.2, it is the $\mathcal{L}$-1-reduct of some model $\langle H', T \rangle$ of Π_1 with $H' \subset T$ such that $H' \cap \mathcal{L} = H \cap \mathcal{L}$. By Lemma 1, since $\langle H, T \rangle \models \Sigma$ also $\langle H', T \rangle \models \Sigma$ and hence $\langle H', T \rangle \models \Pi_1 \cup \Sigma$. But this contradicts the assumption that $\langle T, T \rangle$ is an equilibrium model of $\Pi_1 \cup \Sigma$. Therefore, $\langle T, T \rangle$ is an equilibrium model of $\Pi_2 \cup \Sigma$. □

We now tackle the converse of Theorem 1.

Theorem 2. *Let Π_1 and Π_2 be theories such that Π_1 and Π_2 are strongly equivalent relative to $\mathcal{L}$. Then, they have the same RSE$_\mathcal{L}$-models.*

Proof. Suppose that Π_1 and Π_2 have different RSE$_\mathcal{L}$-models. We shall define a set of $\mathcal{L}$-sentences Σ such that $\Pi_1 \cup \Sigma$ and $\Pi_2 \cup \Sigma$ have different equilibrium models. Without loss of generalisation, assume there is an $\mathcal{M}$ which is an RSE$_\mathcal{L}$-model of Π_1 but not of Π_2. We consider several cases and subcases.

CASE 1. $\mathcal{M} = \langle T, T \rangle$ is a total RSE$_\mathcal{L}$-model of Π_1 that is not an RSE$_\mathcal{L}$-model of Π_2. Set $\Sigma = T \cap \mathcal{L}$. Then clearly $\mathcal{M} \models \Pi_1 \cup \Sigma$. Moreover, $\mathcal{M}$ is an equilibrium model of $\Pi_1 \cup \Sigma$. For, if not, there is a model $\langle H, T \rangle$ of $\Pi_1 \cup \Sigma$ with $H \subset T$. Since $\Sigma = T \cap \mathcal{L}$, we must have $T \cap \mathcal{L} \subseteq H$. But then $T \cap \mathcal{L} = H \cap \mathcal{L}$, which contradicts Condition 4.3 for $\mathcal{M}$ being an RSE$_\mathcal{L}$-model of Π_1. There are two reasons why $\mathcal{M}$ is not an RSE$_\mathcal{L}$-model of Π_2.

SUBCASE 1.1. $\mathcal{M} \not\models \Pi_2$. In this case, since $\mathcal{M} \not\models \Pi_2$, it cannot be an equilibrium model of $\Pi_2 \cup \Sigma$.

SUBCASE 1.2. $\mathcal{M} \models \Pi_2$, but Condition 4.3 fails for Π_2. So, there is a model $\langle J, T \rangle$ of Π_2 with $J \subset T$ such that $T \cap \mathcal{L} = J \cap \mathcal{L}$. Applying Lemma 1, we conclude that $\langle J, T \rangle \models \Sigma$ since $\langle T, T \rangle \models \Sigma$. Therefore, $\langle J, T \rangle \models \Pi_2 \cup \Sigma$, so $\mathcal{M}$ is not an equilibrium model of $\Pi_2 \cup \Sigma$.

CASE 2. $\mathcal{M} = \langle H, T \rangle$ is a non-total RSE$_\mathcal{L}$-model of Π_1 that is not an RSE$_\mathcal{L}$-model of Π_2. Observe that $\langle T, T \rangle$ is a total RSE$_\mathcal{L}$-model of Π_1. Hence, in case $\langle T, T \rangle$ is not an RSE$_\mathcal{L}$-model of Π_2, we can apply the same argument of Case 1 to conclude that $\langle T, T \rangle$ is an equilibrium model of $\Pi_1 \cup \Sigma$ and, again, cannot be an equilibrium model of $\Pi_2 \cup \Sigma$.

So suppose $\langle T, T \rangle$ is an RSE$_\mathcal{L}$-model of Π_2 and Condition 4.2 fails for $\mathcal{M} = \langle H, T \rangle$, i.e., there is no non-total model of Π_2 whose $\mathcal{L}$-1-reduct equals $\mathcal{M}$. Let $\Gamma = \{A \rightarrow B \mid A, B \in (T \setminus H) \cap \mathcal{L}\}$. By Condition 4.3, Γ is non-empty. Set $\Sigma = H \cup \Gamma$. Now, evidently $\langle T, T \rangle$ is a model of both H, since $H \subseteq T$, and of Γ, so $\langle T, T \rangle \models \Pi_2 \cup \Sigma$. We claim it is an equilibrium model of $\Pi_2 \cup \Sigma$, For, if not, there is a model $\langle J, T \rangle$ of $\Pi_2 \cup \Sigma$ with $J \subset T$. Clearly, $H \subseteq J$, but $H \neq J \cap \mathcal{L}$, otherwise $\langle J \cap \mathcal{L}, T \rangle = \mathcal{M}$ would be an RSE$_\mathcal{L}$-model of Π_2. So, $H \subset J \cap \mathcal{L}$. Thus, $(J \cap \mathcal{L}) \setminus H$ is non-empty, and by Condition 4.3, $(T \setminus J) \cap \mathcal{L}$ is also non-empty. Choose an A from $(J \cap \mathcal{L}) \setminus H$ and B from $(T \setminus J) \cap \mathcal{L}$. Then, $A \rightarrow B \in \Gamma$, but $\langle J, T \rangle \not\models A \rightarrow B$, since $\langle J, T \rangle, h \models A$ but $\langle J, T \rangle, h \not\models B$. It follows that $\langle J, T \rangle \not\models \Sigma$ and so $\langle T, T \rangle$ is an equilibrium model of $\Pi_2 \cup \Sigma$. On the other hand, it is not an equilibrium model of $\Pi_1 \cup \Sigma$. In particular, we know that $\langle H', T \rangle \models \Pi_1 \cup H$, since there is a non-total model $\langle H', T \rangle$ of Π_1 whose $\mathcal{L}$-1-reduct equals $\mathcal{M}$. Moreover, $\langle H', T \rangle \models \Gamma$ since $\langle H', T \rangle, h \not\models A$ for each $A \rightarrow B \in \Gamma$ and $\langle H', T \rangle, t \models B$ for each $A \rightarrow B \in \Gamma$. □

4.2 Relativised Uniform Equivalence

We now turn to the characterisation of relativised uniform equivalence via the concept of a relativised uniform equivalence model. First, we mention the following lemma that will be useful later.

Lemma 2. *Suppose Π_1 and Π_2 are theories which are uniformly equivalent relative to $\mathcal{L}$. Then, they have same total RSE$_{\mathcal{L}}$-models.*

Proof. Assume the hypothesis. Suppose Π_1 has a total RSE$_{\mathcal{L}}$-model $\langle T, T\rangle$ that is not a total RSE$_{\mathcal{L}}$-model of Π_2. Evidently, $\langle T, T\rangle \models \Pi_1 \cup (T \cap \mathcal{L})$. Moreover, by Condition 4.3, $\langle T, T\rangle$ must be an equilibrium model of $\Pi_1 \cup (T \cap \mathcal{L})$ since there is no model $\langle J, T\rangle$ of Π_1 with $J \subset T$ such that $T \cap \mathcal{L} \subseteq J \cap \mathcal{L}$. Clearly, if $\langle T, T\rangle \not\models \Pi_2$, it cannot be an equilibrium model of $\Pi_2 \cup (T \cap \mathcal{L})$. On the other hand, if $\langle T, T\rangle \models \Pi_2$ and it is not an RSE$_{\mathcal{L}}$-model of Π_2, then Condition 4.3 fails for Π_2. So, there is a model $\langle J, T\rangle$ of Π_2 with $J \subset T$ such that $T \cap \mathcal{L} = J \cap \mathcal{L}$, whence clearly $\langle T, T\rangle$ is not in equilibrium for $\Pi_2 \cup (T \cap \mathcal{L})$. This contradicts the assumption of relativised uniform equivalence. □

From now on we assume that all theories are finite. As mentioned previously, this means that, under the $\trianglelefteq$-ordering among their models, maximal elements are guaranteed to exist. So, the following notion is well-defined.

Definition 5. *Let Π be a theory in $\mathcal{L}'$ and $\mathcal{L}$ a sublanguage of $\mathcal{L}'$. An RSE$_{\mathcal{L}}$-model of Π is an* RUE$_{\mathcal{L}}$-model *of Π if it is either total or maximal under $\trianglelefteq$ among all non-total RSE$_{\mathcal{L}}$-models of Π.*

Theorem 3. *Let Π_1 and Π_2 be theories which are uniformly equivalent relative to $\mathcal{L}$. Then, they have the same RUE$_{\mathcal{L}}$-models.*

Proof. Assume the hypothesis. By Lemma 2, the two theories have the same total RSE$_{\mathcal{L}}$-models, hence total RUE$_{\mathcal{L}}$-models. Suppose that they differ on non-total RUE$_{\mathcal{L}}$-models, say that Π_1 has a non-total RUE$_{\mathcal{L}}$-model $\langle H, T\rangle$ that is not an RUE$_{\mathcal{L}}$-model of Π_2.

CASE 1. Suppose there is a non-total RSE$_{\mathcal{L}}$-model $\langle J, T\rangle$ of Π_2 with $H \subset J$. So, Π_2 has a non-total model $\langle H', T\rangle$ with $H' \cap \mathcal{L} = J$. Choose an element A from $J \setminus H$ and set $X = H \cup \{A\}$. Clearly, $\langle T, T\rangle \models \Pi_1 \cup X$ and by maximality, $\langle T, T\rangle$ is an equilibrium model of $\Pi_1 \cup X$. On the other hand, by inspection, $\langle H', T\rangle$ is a non-total model of $\Pi_2 \cup X$, so $\langle T, T\rangle$ is not an equilibrium model of $\Pi_2 \cup X$.

CASE 2. Suppose there is no non-total RSE$_{\mathcal{L}}$-model $\langle J, T\rangle$ of Π_2 with $H \subset J$. Since $\langle H, T\rangle$ is not an RUE$_{\mathcal{L}}$-model of Π_2, it cannot be an RSE$_{\mathcal{L}}$-model of Π_2 as well. Consider the model $\langle T, T\rangle$. Since Condition 4.3 holds for Π_1, clearly $\langle T, T\rangle$ is an RSE$_{\mathcal{L}}$-model of Π_1, and hence by Lemma 2 an RSE$_{\mathcal{L}}$-model of Π_2. So, $\langle T, T\rangle \models \Pi_2 \cup H$. Since there is no $H_2 \supseteq H$ such that $H_2 \subset T$ and $\langle H_2, T\rangle \models \Pi_2$, $\langle T, T\rangle$ is an equilibrium model of $\Pi_2 \cup H$. On the other hand, $\langle T, T\rangle$ is not an equilibrium model of $\Pi_2 \cup H$ since $\langle H', T\rangle \models \Pi_1 \cup H$, for some $H' \cap \mathcal{L} = H$. □

Theorem 4. *Suppose that Π_1 and Π_2 are theories with the same RUE$_{\mathcal{L}}$-models. Then, they are uniformly equivalent relative to $\mathcal{L}$.*

Proof. Assume the hypothesis and suppose that for some set X of $\mathcal{L}$ atoms, $\Pi_1 \cup X$ has an equilibrium model $\langle T, T\rangle$ that is not an equilibrium model of $\Pi_2 \cup X$. Clearly, $\langle T, T\rangle$ is a total RUE$_{\mathcal{L}}$-model of Π_1 and so, by assumption, also of Π_2. Therefore, $\langle T, T\rangle \models$

Π_2. Since it is not an equilibrium model of $\Pi_2 \cup X$, there is a model $\langle H, T\rangle \models \Pi_2 \cup X$ with $H \subset T$ and clearly $X \subseteq H$. Then, $\langle H \cap \mathcal{L}, T\rangle$ is an RSE$_\mathcal{L}$-model of Π_2. Keeping T fixed, we extend this to a maximal non-total RSE$_\mathcal{L}$-model $\langle H_2, T\rangle$ of Π_2, where $H \subseteq H_2$. Then, there is a model $\langle H', T\rangle$ of Π_2 such that

$$H' \cap \mathcal{L} = H_2 \supseteq H \cap \mathcal{L}.$$

Evidently, $\langle H_2, T\rangle$ is an RUE$_\mathcal{L}$-model of Π_2. However, it is not even an RSE$_\mathcal{L}$-model of Π_1. If it were, there would be a model $\langle H_1, T\rangle$ of Π_1 with $H_1 \cap \mathcal{L} = H_2$. Since $X \subseteq H \cap \mathcal{L} \subseteq H_1$, $\langle H_1, T\rangle$ would be a non-total model of $\Pi_1 \cup X$, which is impossible by the initial assumption that $\langle T, T\rangle$ is an equilibrium model of $\Pi_1 \cup X$. □

As we have seen in Lemma 2, total RUE$_\mathcal{L}$-models and total RSE$_\mathcal{L}$-models coincide. For non-total RUE$_\mathcal{L}$-models, we obtain an alternative characterisation as follows:

Lemma 3. *Let Π be a theory in $\mathcal{L}'$ and $\mathcal{L}$ a sublanguage of $\mathcal{L}'$. A pair $\langle H, T\rangle$ is a non-total RUE$_\mathcal{L}$-model of Π iff $\langle T, T\rangle \models \Pi$ and it is the $\mathcal{L}$-projection of an* (*unrelativised*) *UE-model $\langle H', T\rangle$ of Π with $H' \cap \mathcal{L} \subset T \cap \mathcal{L}$.*

5 An Application to Prediction and Explanation

In this section, we illustrate how the concept of relativised uniform equivalence can be applied in contexts such as prediction and abductive inference and explanation. Different types of scenarios are possible. For instance, in predicting the behaviour of physical systems we might have a general theory Π comprising strict laws as well as nonmonotonic rules, e.g., describing inertia axioms, default conditions etc., together with initial conditions represented by atomic formulas in a suitable subset of the language. Another type of scenario is represented by an *abductive logic program*, $\langle \Pi, \mathcal{A}\rangle$, where Π is a logic program (of any general type, e.g., disjunctive, nested, etc.) and $\mathcal{A}$ is a set of literals called *abducibles* in a suitable sublanguage of Π. In each case, we are interested in the question: When are two such "theories" equivalent in terms of predictive power, explanatory capacity, and so on? The structure of inference is similar in the two cases mentioned. In each case, the theory Π conjoined with a set $\{A_1, \ldots, A_n\}$ of literals representing initial conditions, abducibles, etc., entails a sentence, say φ, representing, e.g., the prediction of a physical state, the effects of an action, or an explanandum in an abductive system. In the context of equilibrium logic and ASP, entailment is of course nonmonotonic.[2]

To fix notation and terminology, let us consider the general case of *abductive theories*, which are given as pairs of form $\langle \Pi, \mathcal{A}\rangle$, where Π is a theory and $\mathcal{A}$ is a set of literals, and the matter of equivalence with respect to abductive explanations. This leads to the following definition.

[2] The main difference between a prediction in the former sense and an abductive explanation in the latter sense is *methodological*: in the first case, the literals $\{A_1, \ldots, A_n\}$ are specified in advance as part of the initial conditions of the system, while in the second case, it is φ that is supplied in advance as an explanandum, and the abducibles $\{A_1, \ldots, A_n\}$ are to be discovered.

Definition 6. *An* abductive explanation *of a sentence φ by an abductive theory $\mathcal{P} = \langle \Pi, \mathcal{A} \rangle$ is a set $\{A_1, \ldots, A_n\}$ satisfying*

$$\Pi \cup \{A_1, \ldots, A_n\} \mid\!\sim \varphi \tag{1}$$

as well as the following two conditions:

6.1 $\{A_1, \ldots, A_n\} \subseteq \mathcal{A}$ *and*
6.2 $\Pi \cup \{A_1, \ldots, A_n\}$ *is consistent,*

where $\mid\!\sim$ is nonmonotonic entailment.[3]

Note that Condition 6.2 merely ensures that the explanation of φ is non-trivial. For present purposes we do not, however, insist that $\{A_1, \ldots, A_n\}$ be a minimal set of abducibles explaining φ, nor even that it is non-empty. We note further that Definition 6 is equivalent to the definition of an abductive explanation as given by Inoue and Sakama [8] for the case of disjunctive logic programs with default negations in their heads.

If $\{A_1, \ldots, A_n\}$ is an abductive explanation of φ from $\mathcal{P}$, then we also say that $\{A_1, \ldots, A_n\}$ *explains φ in $\mathcal{P}$*. $\mathcal{P}$ is said to have *explanatory power* if there exist some φ and $\{A_1, \ldots, A_n\}$ satisfying (1) as well as Conditions 6.1 and 6.2. Evidently, two abductive theories can have the same explanatory power in weaker or stronger senses. They may capture the same explananda by means of possibly differing explanans (abducibles), and therefore differing explanations, or they may support essentially the same explanations. In this latter sense, we can say therefore that two abductive theories, $\mathcal{P}_1$ and $\mathcal{P}_2$, based on the same abducible set $\mathcal{A}$, have the *same explanatory power in the strong sense* if, for any φ and any $\{A_1, \ldots, A_n\} \subseteq \mathcal{A}$, $\{A_1, \ldots, A_n\}$ explains φ in $\mathcal{P}_1$ iff $\{A_1, \ldots, A_n\}$ explains φ in $\mathcal{P}_2$. We consider here only abductive theories with (non-vacuous) explanatory power.

We can easily relate this notion of explanatory equivalence to relativised uniform equivalence. The following is straightforward.

Proposition 1. *Let $\mathcal{P}_1 = \langle \Pi_1, \mathcal{A} \rangle$ and $\mathcal{P}_2 = \langle \Pi_2, \mathcal{A} \rangle$ be abductive theories based on the same abducibles. If Π_1 and Π_2 are uniformly equivalent relative to $\mathcal{A}$, then $\mathcal{P}_1$ and $\mathcal{P}_2$ have the same explanatory power* (*in the strong sense*).

If Π_1 and Π_2 are uniformly equivalent relative to $\mathcal{A}$, then for any $\{A_1, \ldots, A_n\} \subseteq \mathcal{A}$, $\Pi_1 \cup \{A_1, \ldots, A_n\}$ and $\Pi_2 \cup \{A_1, \ldots, A_n\}$ have the same equilibrium models, so the explanatory power of $\mathcal{P}_1$ and $\mathcal{P}_2$ is the same whether we interpret entailment $\mid\!\sim$ in the cautious or brave sense.

To establish a converse of Proposition 1, we need to pin down the type of inference defined by $\mid\!\sim$. Evidently, brave reasoning has a greater chance of succeeding, since *prima facie* it seems possible that theories might have the same consequences in the cautious sense, even under the addition of new atoms, yet have different equilibrium models and therefore not be relativised uniformly equivalent.

So let us suppose that $\mid\!\sim$ is entailment with respect to to some equilibrium model; in other words, $\Pi \mid\!\sim \varphi$ iff φ is true in some equilibrium model of Π. Then we have:

[3] We leave open for the moment whether entailment is to be understood in the cautious or brave sense.

Proposition 2. *If $\mathcal{P}_1$ and $\mathcal{P}_2$ have the same* (*non-vacuous*) *explanatory power* (*in the strong sense*)*, then Π_1 and Π_2 are uniformly equivalent relative to $\mathcal{A}$.*

Proof. Assume the hypothesis of the proposition and suppose that they are not uniformly equivalent relative to $\mathcal{A}$. Then, there exists a subset $\{A_1, \ldots, A_n\} \subseteq \mathcal{A}$ such that $\Pi_1 \cup \{A_1, \ldots, A_n\}$ and $\Pi_2 \cup \{A_1, \ldots, A_n\}$ have different equilibrium models. Say, $\Pi_1 \cup \{A_1, \ldots, A_n\}$ has an equilibrium model $\mathcal{M}$ that is not an equilibrium model of $\Pi_2 \cup \{A_1, \ldots, A_n\}$. We can establish that $\mathcal{P}_1$ and $\mathcal{P}_2$ have different explanatory powers if we can find a sentence φ that is true in $\mathcal{M}$, so that

$$\Pi_1 \cup \{A_1, \ldots, A_n\} \mid\!\sim \varphi \tag{2}$$

but

$$\Pi_2 \cup \{A_1, \ldots, A_n\} \not\mid\!\sim \varphi. \tag{3}$$

This means that φ has to be chosen so there is no other equilibrium model of $\Pi_2 \cup \{A_1, \ldots, A_n\}$ in which φ is true. Moreover, Conditions 6.1 and 6.2 above should also hold for (2). By assumption, no equilibrium model of $\Pi_2 \cup \{A_1, \ldots, A_n\}$ can be equivalent to $\mathcal{M}$ in that it satisfies exactly the same sentences; otherwise it would make exactly the same literals true and false and so be exactly $\mathcal{M}$. So, for each equilibrium model $\mathcal{M}_i$ of $\Pi_2 \cup \{A_1, \ldots, A_n\}$, there must be some sentence α_i true in $\mathcal{M}$ that is not true in $\mathcal{M}_i$. Since we are assuming that the theories are finite, there are at most finitely many equilibrium models $\mathcal{M}_i$ of $\Pi_2 \cup \{A_1, \ldots, A_n\}$ and therefore finitely many such α_i. Evidently, the sentence $\bigwedge_i \alpha_i$ is true in $\mathcal{M}$ but not true in any equilibrium model of $\Pi_2 \cup \{A_1, \ldots, A_n\}$. So, we have

$$\Pi_1 \cup \{A_1, \ldots, A_n\} \mid\!\sim \bigwedge_i \alpha_i \text{ and} \tag{4}$$

$$\Pi_2 \cup \{A_1, \ldots, A_n\} \not\mid\!\sim \bigwedge_i \alpha_i. \tag{5}$$

Furthermore, we have that 6.1 is satisfied and 6.2 holds since $\Pi_1 \cup \{A_1, \ldots, A_n\}$ has a model. This contradicts the initial assumption that $\mathcal{P}_1$ and $\mathcal{P}_2$ have the same explanatory power. □

Combining Propositions 1 and 2 with Theorems 3 and 4 yields the following semantic characterisation of explanatory equivalence.

Corollary 1. *Two abductive theories $\mathcal{P}_1 = \langle \Pi_1, \mathcal{A} \rangle$ and $\mathcal{P}_2 = \langle \Pi_2, \mathcal{A} \rangle$ have the same explanatory power* (*in the strong sense*) *iff Π_1 and Π_2 have the same RUE$_{\mathcal{A}}$-models.*

We note that Inoue and Sakama [8, 9] provided for the case of abductive logic programs with default negations in the heads a characterisation similar to our Propositions 1 and 2. However, they derived that two abductive programs $\langle \Pi_1, \mathcal{A} \rangle$ and $\langle \Pi_1, \mathcal{A} \rangle$ have the same explanatory power iff Π and Π_2 are strongly equivalent relative to $\mathcal{A}$. In view of our results, it seems that relativised strong equivalence should in their characterisation be replaced by relative uniform equivalence. Because otherwise we would obtain that, for any $\mathcal{A}$, strong equivalence relative to $\mathcal{A}$ would coincide with uniform equivalence relative to $\mathcal{A}$, which is obviously violated (consider, e.g., the programs $\{a \vee b \leftarrow\}$ and $\{a \leftarrow not\, b;\ b \leftarrow not\, a\}$ which are uniformly equivalent relative to $\{a, b\}$ but not strongly equivalent relative to $\{a, b\}$). Let us also note that they do not apply any semantic characterisations of equivalence analogous to Corollary 1 above. On the other hand, they also consider equivalence in the context of an extended abduction concept [9].

6 Complexity

The complexity of relativised equivalence between disjunctive logic programs has been established by Woltran [19] and has been further studied by Eiter, Fink, and Woltran [2]. Both notions, i.e., RSE and RUE, yield Π_2^P-complete decision problems. Thus, Π_2^P-hardness for these problems is immediate for equilibrium logic. To show that RSE and RUE remain in class Π_2^P for the general setting studied here, first observe that the central subtask of checking whether a given pair $\langle T, T\rangle$ is an equilibrium model of some theory Π is in coNP. Moreover, to decide the complementary problem of RUE between Π_1 and Π_2, one can guess sets T, F of literals and check whether $\langle T, T\rangle$ is an equilibrium model of exactly one of $\Pi_1 \cup F$ and $\Pi_2 \cup F$. This algorithm runs in non-deterministic time with access to an NP-oracle, and thus in Σ_2^P. Π_2^P-membership for RUE follows immediately. The same argumentation holds for RSE in view of the proof of Theorem 2, where it is shown that only very simple theories (which are polynomial in the size to the compared programs) are sufficient to decide RSE.

7 Conclusions and Future Work

In this paper, we extended results for relativised notions of equivalence from logic programs under the answer-set semantics to arbitrary (propositional) theories in equilibrium logic. To this end, we introduced the concept of an $\mathcal{L}$-1-reduct which restricts the language of one world in the two-world Kripke-model for equilibrium logic. These partially bound models can be shown to characterise relativised strong and uniform equivalence between theories in the same manner as relativised SE- and UE-models are used for logic programs [19]. Furthermore, we discussed a possible application of relativised equivalences in the area of abduction and we briefly studied the complexity of the introduced equivalence notions.

An interesting topic for further work is to extend our notions to include the removal of auxiliary letters—important for considering submodules of theories having dedicated output atoms—tantamount to considering *projected equilibrium models*, where only a subset of the atoms are of interest. This would be an extension of the framework introduced by Eiter, Fink, and Woltran [3] for disjunctive logic programs.

References

1. T. Eiter and M. Fink. Uniform equivalence of logic programs under the stable model semantics. In *Proc. ICLP 2003*, vol. 2916 in LNCS, pg. 224–238. Springer, 2003.

2. T. Eiter, M. Fink, and S. Woltran. Semantical characterizations and complexity of equivalences in stable logic programming. *ACM Transactions on Computational Logic*, 2007. To appear.
3. T. Eiter, H. Tompits, and S. Woltran: On solution correspondences in answer-set programming. In *Proc. IJCAI 2005*, pg. 97–102, 2005.
4. P. Ferraris and V. Lifschitz. Weight constraints as nested expressions. *Theory and Practice of Logic Programming* 5:45–74, 2005.
5. P. Ferraris. Answer sets for propositional theories. in *Proc. LPNMR 2005*, vol. 3662 in LNCS, pg. 119–131. Springer, 2005.
6. Y. Gurevich. Intuitionistic logic with strong negation. *Studia Logica*, 36(1–2):49–59, 1977.
7. K. Inoue and C. Sakama. Equivalence of logic programs under updates. In *Proc. JELIA 2004*, vol. 3229 of LNCS, pg. 174–186. Springer, 2004.
8. K. Inoue and C. Sakama. Equivalence in abductive logic. In *Proc. IJCAI 2005*, pg. 472–477, 2005.
9. K. Inoue and C. Sakama. On abductive equivalence. In L. Magnani (ed), *Model-Based Reasoning in Science and Engineering. Cognitive Science, Epistemology, Logic*, Studies in Logic, pg. 333–352, College Publications, London, 2006.
10. M. Kracht. On extensions of intermediate logics by strong negation. *Journal of Philosophical Logic*, 27(1):49–73, 1998.
11. V. Lifschitz, D. Pearce, and A. Valverde. Strongly equivalent logic programs. *ACM Transactions on Computational Logic*, 2(4):526–541, 2001.
12. D. Pearce. A new logical characterisation of stable models and answer sets. In *Non-Monotonic Extensions of Logic Programming* (*NMELP'96*), vol. 1216 of LNCS, pg. 57–70. Springer, 1997.
13. D. Pearce. Equilibrium logic. *Annals of Mathematics and Artificial Intelligence*, 47(1-2):3–41, 2006.
14. D. Pearce and A. Valverde. Uniform equivalence for equilibrium logic and logic programs. In *Proc. LPNMR 2004*, vol. 2923 of LNCS, pg. 194–206. Springer, 2004.
15. H. Turner. Strong equivalence for logic programs and default theories (made easy). In *Proc. LPNMR 2001*, vol. 2173 of LNCS, pg. 81–92. Springer, 2001.
16. D. van Dalen. Intuitionistic logic. In *Handbook of Philosophical Logic, Volume III: Alternatives in Classical Logic*. D. Reidel Publishing, 1986.
17. N. N. Vorob'ev. A constructive propositional calculus with strong negation (in Russian). *Doklady Akademii Nauk SSR*, 85:465–468, 1952.
18. N. N. Vorob'ev. The problem of deducibility in constructive propositional calculus with strong negation (in Russian). *Doklady Akademii Nauk SSR*, 85:689–692, 1952.
19. S. Woltran. Characterizations for relativized notions of equivalence in answer set programming. in *Proc. JELIA 2004*, vol. 3229 of LNCS, pg. 161–173. Springer, 2004.