# A REVISED CONCEPT OF SAFETY FOR GENERAL ANSWER SET PROGRAMS

Pedro Cabalar, David Pearce and Agustín Valverde

ELWG TECHNICAL REPORT 2009-01
APRIL 2009, REVISED OCTOBER 2009

**Equilibrium Logic Working Group**
ESPAÑA
http://www.equilibriumlogic.net/

ELWG TECHNICAL REPORT

ELWG TECHNICAL REPORT 2009-01, APRIL 2009, REVISED OCTOBER 2009

# A REVISED CONCEPT OF SAFETY FOR GENERAL ANSWER SET PROGRAMS (EXTENDED VERSION)

**Pedro Cabalar**
Department of Computer Science,
University of Corunna, Spain.
cabalar@udc.es

**David Pearce**
Department of Artficial Intelligence,
Universidad Politécnica de Madrid, Spain.
david.pearce@upm.es

**Agustín Valverde**
Department of Applied Mathematics,
Universidad de Málaga, Spain.
avalverde@ctima.uma.es

**Abstract.**Some answer set solvers deal with programs with variables by requiring a safety condition on program rules. This ensures that there is a close relation between the answer sets of the program and those of its ground version. If we move beyond the syntax of disjunctive programs, for instance by allowing rules with nested expressions, or perhaps even arbitrary first-order formulas, new definitions of safety are required. In this paper we consider a new concept of safety for formulas in quantified equilibrium logic where answer sets can be defined for arbitrary first-order formulas. The new concept captures and generalises two recently proposed safety definitions: that of Lee, Lifschitz and Palla (2008) as well as that of Bria, Leone and Faber (2008). We study the main metalogical properties of safe formulas.

**Acknowledgements**: We gratefully acknowledge support from the Spanish MEC (now MICINN) under the projects TIN2006-15455-(C01,C02,C03) and Agreement Technologies, Consolider CSD2007-00022, and the Junta de Andalucia, project P6-FQM-02049.

## 1 Introduction

In answer set programming (ASP) more and more tools are being created to enable a fuller use of first-order languages and logic. Recent developments include efforts to extend the syntax of programs as well as to deal more directly with variables in a full, first-order context. In several cases, assumptions such as standard names (SNA) are being relaxed and issues involving programming in open domains are being addressed.

A stable model semantics for first-order structures and languages was defined in the framework of equilibrium logic in [16,17,18]. As a logical basis the non-classical logic of quantified here-and-there, **QHT**, is used (see also [14]). By expanding the language to include new predicates, this logic can be embedded in classical first-order logic, [19], and this permits an alternative but equivalent formulation of the concept of stable model for first-order formulas, expressed in terms of classical, second-order logic [6]. The latter definition of answer set has been further studied in [10,11] where the basis of a first-oder programming language, RASPL-1, is described. An alternative approach to a first-order ASP language is developed in [8,7].

Several implementations of answer set programming deal with programs with variables by requiring a safety condition on program rules. In the language of disjunctive LPs, this condition is usually expressed by saying that a rule is safe if any variable in the rule also appears in its positive body – this condition will be referred here as *DLP safety*. Programs are safe if all their rules are safe. The safety of a program ensures that its answer sets coincide with the answer sets of its ground version and thus allows ASP systems to be based on computations at the level of propositional logic which may include for example the use of SAT-solvers.

What happens if we move beyond the syntax of disjunctive programs? Adding negation in the heads of program rules will not require a change in the definition of safety. But for more far reaching language extensions, such as allowing rules with nested expressions, or perhaps even arbitrary first-order formulas, new definitions of safety are required. Once again, the desideratum is that safe programs should be in a certain sense reducible to their ground versions, whether or not these can be compiled into an existing ASP-solver.

Recently, there have been two new extensions of the notion of safety to accommodate different syntactic extensions of the ASP language in the first-order case. In [2], Bria, Faber and Leone introduce an extension to a restricted variant of the language of programs with nested expressions (or nested programs, for short) [15]. The rules are called *normal form nested* or NFN rules. The safety of such rules which will be referred here as *BFL-safety* is designed to guarantee domain independence and permit their efficient translation into the language of disjunctive programs. In this manner, NFN programs can be compiled into the standard DLV system. A second example is the definition of safety introduced by Lee, Lifschitz and Palla in [11] (we will call it here *LLP-safety*) defined essentially for arbitrary first-order formulas in such a way that the stable model of such a formula can be obtained by considering its ground version. It is also shown that the stable models of safe formulas form a (first-order) definable class. Here the motivation is not so much to compile safe programs into existing ASP languages, but rather to find tractable, first-order extensions of the customary languages.[1]

On the face of it, the approach followed in [11] is much more general than that of [2] and so we might expect that the safety of general formulas from [11] embraces also the safe NFN rules of [2]. But this is not the case: safe NFN rules need not be safe formulas in the sense of [11]. In fact, there are simple examples of formulas that would normally be regarded as safe that fail to be so under the [11] definition. Consider the following example.

$$p \leftarrow q(X) \vee r \tag{1}$$

This rule belongs to the syntactic class of programs with nested expressions as introduced in [15], although (1) contains a variable, something not considered in that work. Still, this rule is strongly equivalent to the conjunction of rules

$$p \leftarrow q(X)$$
$$p \leftarrow r$$

and both are DLP safe (note that this is now a normal logic program) and so, they are BFL-safe and LLP-safe as well. Therefore, it is reasonable to consider (1) as safe. According to [11], however, for (1) to be safe, $X$ must occur in some

[1] Note that in *open* answer set programming an alternative approach to decidable language extensions is adopted. Rather than generalise safety conditions, this approach defines guarded fragments [7].

implication $G \rightarrow H$ such that $X$ belongs to what are called the restricted variables of $G$. In this case, the only option for $G$ is the rule body $q(X) \vee r$ whose set of restricted variables is empty, implying that the rule is not safe. [2] also uses a similar concept of restricted variable but only imposes restrictions on variables occurring in the head of the rule or within the scope of negation; so according to this account, (1) is safe.

In this paper we propose a new definition of safety that, while defined for arbitrary first order formulas, is strictly weaker than LLP-safety, while it captures BFL-safety when restricted to the same syntactic class of NFN-programs.

**desiderata for safety** For our purposes there are three key properties associated with safety that we should keep distinct. The first property is very simple and does not refer to grounding. It says merely that a stable model should not contain unnamed individuals. This condition is formally expressed by Theorem 2 below and is fulfilled by formulas that we call *semi-safe*. Secondly we have the property usually most associated with safety. This is called *domain independence* and it does refer to grounding. It says that grounding a program with respect to any superset of the program's constants will not change the class of stable models. This is satisfied by formulas that we call *safe* and is expressed by Proposition 3 and Theorem 4. The third property, also satisfied by safe formulas, is expressed in Theorem 5: it says that the class of stable models should be first-order definable. This may be relevant for establishing properties such as interpolation and for computational purposes, being exploited for instance by the method of loop formulas.

It is perhaps interesting to consider some inherent limitations of a safety definition. Arbitrary theories are, in general, not domain independent. Even simple normal rules like, for instance:

$$p \leftarrow \mathit{not}\ q(X) \tag{2}$$

are not domain independent. To see why, take (2) plus the fact $q(1)$. If we ground this program on the domain $\{1\}$ we get the stable model $\{q(1)\}$, but if we use the extended domain $\{1, 2\}$, the only stable model becomes now $\{q(1), p\}$. Unfortunately, directly checking domain independence of an arbitrary formula does not seem a trivial task. It seems much more practical to find a (computationally) simple syntactic condition, like safety, that suffices to guarantee domain independence. However, due to its syntactic nature, a condition of this kind will just be *sufficient* but, most probably, not *necessary*. For instance, while as we saw above, (2) cannot be considered safe under any definition, if we consider the conjunction of (2) with the (variable-free) constraint:

$$\bot \leftarrow q(1) \tag{3}$$

the resulting program is *domain independent*. The reason is that (2) is strongly equivalent to the formula $(\exists x\ \neg q(x)) \rightarrow p$ whereas (3) is equivalent to $\neg q(1)$ that, in its turn, implies $(\exists x\ \neg q(x))$ in $\mathbf{QHT}$. This example shows that checking domain independence may require considering semantic interrelations of the formula or program as a whole. On the other hand, the fact that safety is sensitive to redundancies or that the safety of a formula may vary under a strongly equivalent transformation is unavoidable. For instance, the program consisting of (2) and the fact $p$ is clearly domain independent, as in the presence of $p$, (2) becomes redundant and can be removed.

As regards our methodology, it is important to note that in order to handle explicit quantifiers and possible relaxations of the SNA it is indispensable to use a genuine first-order (or higher-order) semantics. Our approach here will be to analyse safety from the standpoint of (quantified) equilibrium logic in which a first-order definition of stable or equilibrium model can be given in the logic $\mathbf{QHT}$. One specific aim is to generalise the safety concept from [11] so that rules such as (1), and indeed all the safe rules from [2], can be correctly categorised as safe. At least as regards simplicity and intuitiveness, we feel that our approach has some advantages over the methodology of [11] that uses a mixture of $\mathbf{QHT}$ and classical second-order logic.

## 2 Review of Quantified Equilibrium Logic and Answer Sets

Usually in quantified equilibrium logic we consider a full first-order language allowing function symbols and we include a second, strong negation operator as occurs in several ASP dialects. In this paper we shall restrict attention to the function-free language with a single negation symbol, '$\neg$'. In particular, we shall work with a quantified version of the logic $\mathbf{HT}$ of *here-and-there*. In other respects we follow the treatment of [18].

For the remainder of the paper we consider function-free first order languages $\mathcal{L} = \langle C, P\rangle$ built over a set of *constant* symbols, $C$, and a set of *predicate* symbols, $P$. The sets of $\mathcal{L}$-formulas, $\mathcal{L}$-sentences and atomic $\mathcal{L}$-sentences are defined in the usual way. We work here mainly with *sentences*. If $D$ is a non-empty set, we denote by $At(D, P)$ the set of ground atomic sentences of $\langle D, P\rangle$. By an $\mathcal{L}$-interpretation $I$ over a set $D$ we mean a subset of $At(D, P)$. A *classical* $\mathcal{L}$-structure can be regarded as a tuple $\mathcal{M} = \langle (D, \sigma), I\rangle$ where $I$ is an $\mathcal{L}$-interpretation over $D$ and $\sigma\colon C \cup D \to D$ is a mapping, called the *assignment*, such that $\sigma(d) = d$ for all $d \in D$. If $D = C$ and $\sigma = id$, $\mathcal{M}$ is a *Herbrand structure*.

On the other hand, a *here-and-there* $\mathcal{L}$-structure with static domains, or $\mathbf{QHT}(\mathcal{L})$*-structure*, is a tuple $\mathcal{M} = \langle (D, \sigma), I_h, I_t\rangle$ where $\langle (D, \sigma), I_h\rangle$ and $\langle (D, \sigma), I_t\rangle$ are classical $\mathcal{L}$-structures such that $I_h \subseteq I_t$.

Thus we can think of a here-and-there structure $\mathcal{M}$ as similar to a first-order classical model, but having two parts, or components, $h$ and $t$ that correspond to two different points or "worlds", 'here' and 'there', in the sense of Kripke semantics for intuitionistic logic [20], where the worlds are ordered by $h \leq t$. At each world $w \in \{h, t\}$ one verifies a set of atoms $I_w$ in the expanded language for the domain $D$. We call the model static, since, in contrast to say intuitionistic logic, the same domain serves each of the worlds. Since $h \leq t$, whatever is verified at $h$ remains true at $t$. The satisfaction relation for $\mathcal{M}$ is defined so as to reflect the two different components, so we write $\mathcal{M}, w \models \varphi$ to denote that $\varphi$ is true in $\mathcal{M}$ with respect to the $w$ component. The recursive definition of the satisfaction relation, forces us to consider formulas from $\langle C \cup D, P\rangle$. Evidently we should require that an atomic sentence is true at $w$ just in case it belongs to the $w$-interpretation. Formally, if $p(t_1, \ldots, t_n) \in At(C \cup D, P)$ and $w \in \{h, t\}$ then

$$\begin{aligned} \mathcal{M}, w \models p(t_1, \ldots, t_n) \quad &\text{iff} \quad p(\sigma(t_1), \ldots, \sigma(t_n)) \in I_w. \\ \mathcal{M}, w \models t = s \quad &\text{iff} \quad \sigma(t) = \sigma(s) \end{aligned}$$

Then $\models$ is extended recursively as follows, conforming to the usual Kripke semantics for intuitionistic logic given our assumptions about the two worlds $h$ and $t$ and the single domain $D$, see eg. [20]. We shall assume that $\mathcal{L}$ contains the constants $\top$ and $\bot$ and regard $\neg\varphi$ as an abbreviation for $\varphi \to \bot$.

- $\mathcal{M}, w \models \top,\ \mathcal{M}, w \not\models \bot$
- $\mathcal{M}, w \models \varphi \wedge \psi$ iff $\mathcal{M}, w \models \varphi$ and $\mathcal{M}, w \models \psi$.
- $\mathcal{M}, w \models \varphi \vee \psi$ iff $\mathcal{M}, w \models \varphi$ or $\mathcal{M}, w \models \psi$.
- $\mathcal{M}, t \models \varphi \to \psi$ iff $\mathcal{M}, t \not\models \varphi$ or $\mathcal{M}, t \models \psi$.
- $\mathcal{M}, h \models \varphi \to \psi$ iff $\mathcal{M}, t \models \varphi \to \psi$ and $\mathcal{M}, h \not\models \varphi$ or $\mathcal{M}, h \models \psi$.
- $\mathcal{M}, t \models \forall x\varphi(x)$ iff $\mathcal{M}, t \models \varphi(d)$ for all $d \in D$.
- $\mathcal{M}, h \models \forall x\varphi(x)$ iff $\mathcal{M}, t \models \forall x\varphi(x)$ and $\mathcal{M}, h \models \varphi(d)$ for all $d \in D$.
- $\mathcal{M}, w \models \exists x\varphi(x)$ iff $\mathcal{M}, w \models \varphi(d)$ for some $d \in D$.

Truth of a sentence in a model is defined as follows: $\mathcal{M} \models \varphi$ iff $\mathcal{M}, w \models \varphi$ for each $w \in \{h, t\}$. A sentence $\varphi$ is a consequence of a set of sentences $\Gamma$, denoted $\Gamma \models \varphi$, if every model of $\Gamma$ is a model of $\varphi$. In a model $\mathcal{M}$ we also use the symbols $H$ and $T$, possibly with subscripts, to denote the interpretations $I_h$ and $I_t$ respectively; so, an $\mathcal{L}$-structure may be written in the form $\langle U, H, T\rangle$, where $U = (D, \sigma)$. A structure $\langle U, H, T\rangle$ is called *total* if $H = T$, whence it is equivalent to a classical structure.

We shall also make use of an equivalent semantics based on many-valued logic. In the propositional case the Kripke semantics is easily characterised using a three-valued matrix: the set of truth values is $\mathbf{3} = \{0, 1, 2\}$ and 2 is the designated value; the connectives are interpreted as follows: $\wedge$ is the minimum function, $\vee$ is the maximum function,

$$x \to y = \begin{cases} 2 & \text{if either } x = 0 \text{ or } x \leq y \\ y & \text{otherwise} \end{cases} \qquad \text{and} \qquad \neg x = \begin{cases} 2 & \text{if } x = 0 \\ 0 & \text{otherwise} \end{cases}$$

If we add quantifiers using the standard approach for many-valued logics, we obtain a semantics which we can denote by $\mathbf{QN}_3$ as follows: an interpretation is an assignment $I\colon At(D, P) \to \mathbf{3}$ extended to any sentence with the previous rules and

$$\sigma(\forall x\varphi(x)) = \min\{\sigma(\varphi(d)); d \in D\} \qquad \sigma(\exists x\varphi(x)) = \max\{\sigma(\varphi(d)); t \in D\}$$

In the propositional case a three-valued $\sigma$ as a truth-value assignment can trivially be converted into a Kripke model $\langle H, T\rangle$, and *vice versa*. In the first-order case the Kripke semantics with static domains and the many-valued semantics are also equivalent.

**Theorem 1.** *There is a bijection $f$ between* $\mathbf{QHT}$*-models and* $\mathbf{QN}_3$*-models such that for any formula $\varphi$ and* $\mathbf{QHT}$*-model $\mathcal{M}$, $\mathcal{M} \models \varphi$ iff $f(\mathcal{M})(\varphi) = 2$.*

$$\begin{array}{lll} \sigma(\varphi) = 2 & \text{iff} & \mathcal{M}, h \models \varphi \\ \sigma(\varphi) = 1 & \text{iff} & \mathcal{M}, t \models \varphi, \mathcal{M}, h \not\models \varphi \\ \sigma(\varphi) = 0 & \text{iff} & \mathcal{M}, t \not\models \varphi \end{array}$$

In other words we can equivalently work with Kripke models having static domains or with the five-valued semantics. In the sequel we use both semantics.

The resulting logic is called *Quantified Here-and-There Logic with static domains and decidable equality*, and denoted in [14] by $\mathbf{SQHT}^{=}$. In terms of satisfiability and validity this logic is equivalent to the logic previously introduced in [17]. To simplify notation we drop the labels for static domains and equality and refer to this logic simply as quantified here-and-there, $\mathbf{QHT}$.

A complete axiomatisation of $\mathbf{QHT}$ can be obtained as follows [14]. We take the axioms and rules of first-order intuitionistic logic [20] and add the axiom of Hosoi

$$\alpha \vee (\neg\beta \vee (\alpha \rightarrow \beta))$$

which determines 2-element here-and-there models in the propositional case, together with two further axioms, the second of which captures decidable equality:

$$\exists x(\alpha(x) \rightarrow \forall x\alpha(x))$$
$$\forall x \forall y((x = y) \vee \neg(x = y))$$

In the context of logic programs, the following assumptions often play a role. Let $\sigma|_C$ denote the restriction of the assignment $\sigma$ to constants in $C$. In the case of both classical and $\mathbf{QHT}$ models, we say that the *parameter names assumption* (PNA) applies in case $\sigma|_C$ is surjective, i.e., there are no unnamed individuals in $D$; the *unique names assumption* (UNA) applies in case $\sigma|_C$ is injective; in case both the PNA and UNA apply, the *standard names assumption* (SNA) applies, i.e. $\sigma|_C$ is a bijection. A Herbrand structure is an SNA-model in which $\sigma = \mathrm{id}$. In the following, we will speak about PNA-, UNA-, or SNA-models, respectively, depending on $\sigma$.

As in the propositional case, quantified equilibrium logic, or QEL, is based on a suitable notion of minimal model.

**Definition 1.** *Let $\varphi$ be an $\mathcal{L}$-sentence. An equilibrium model of $\varphi$ is a total model $\mathcal{M} = \langle (D, \sigma), T, T\rangle$ of $\varphi$ such that there is no model of $\varphi$ of the form $\langle (D, \sigma), H, T\rangle$ where $H$ is a proper subset of $T$.*

### 2.1 relation to answer sets

We assume the reader is familiar with the usual definitions of answer set based on *Herbrand* models and ground programs, eg. [1]. Two variations of this semantics, the open [8] and generalised open answer set [9] semantics, consider non-ground programs and open domains, thereby relaxing the PNA.

For the present version of QEL the correspondence to answer sets can be summarised as follows (see [17,18,3]). If $\varphi$ is a universal sentence in $\mathcal{L} = \langle C, P\rangle$ (see §3 below), a total $\mathbf{QHT}$ model $\langle U, T, T\rangle$ of $\varphi$ is an equilibrium model of $\varphi$ iff $\langle T, T\rangle$ is a propositional equilibrium model of the grounding of $\varphi$ with respect to the universe $U$.

By the usual convention, when $\Pi$ is a logic program with variables we consider the models and equilibrium models of its universal closure expressed as a set of logical formulas. It follows that if $\Pi$ is a logic program (of any form), a total $\mathbf{QHT}$ model $\langle U, T, T\rangle$ of $\Pi$ is an equilibrium model of $\Pi$ iff it is a generalised open answer set of $\Pi$ in the sense of [9]. If we assume all models are UNA-models, we obtain the version of QEL found in [17]. There, the following relation of QEL to (ordinary) answer sets for logic programs with variables was established. If $\Pi$ is a logic program, a total UNA-$\mathbf{QHT}$ model $\langle U, T, T\rangle$ of $\Pi$ is an equilibrium model of $\Pi$ iff it is an open answer set of $\Pi$.

In [6] Ferraris, Lee, and Lifschitz provide a new definition of stable model for arbitrary first-order formulas. In this case the property of being a stable model is defined syntactically via a second-order condition. However in [6] it is also shown that the new notion of stable model is equivalent to that of equilibrium model defined here. In a sequel

to this paper, Lee, Lifschitz and Palla [10] have applied the new definition and made the following refinements. The *stable models* of a formula are defined as in [6] while the *answer sets* of a formula are those Herbrand models of the formula that are stable in the sense of [6]. Using this new terminology, it follows that in general stable models and equilibrium models coincide, while answer sets are equivalent to SNA-**QHT** models that are equilibrium models.

## 3 Safety

We work with a concept of restricted variable that was introduced in [10,11]. To every quantifier-free formula $\varphi$ the set $\mathrm{RV}(\varphi)$ of its *restricted variables* is defined as follows:

- For an atomic formula $\varphi$,
  - if $\varphi$ is an equality between two variables then $\mathrm{RV}(\varphi) = \varnothing$;
  - otherwise, $\mathrm{RV}(\varphi)$ is the set of all variables occurring in $\varphi$;
- $\mathrm{RV}(\bot) = \varnothing$;
- $\mathrm{RV}(\varphi_1 \wedge \varphi_2) = \mathrm{RV}(\varphi_1) \cup \mathrm{RV}(\varphi_2)$;
- $\mathrm{RV}(\varphi_1 \vee \varphi_2) = \mathrm{RV}(\varphi_1) \cap \mathrm{RV}(\varphi_2)$;
- $\mathrm{RV}(\varphi_1 \rightarrow \varphi_2) = \varnothing$.

A sentence is said to be in *prenex* form if it has the following shape, for some $n \geq 0$:

$$Q_1 x_1 \ldots Q_n x_n \psi \tag{4}$$

where $Q_i$ is $\forall$ or $\exists$ and $\alpha$ is quantifier-free. A sentence is said to be *universal* if it is in prenex form and all quantifiers are universal. A universal theory is a set of universal sentences. For **QHT**, normal forms such as prenex and Skolem forms were studied in [17]. In particular it is shown there that in quantified here-and-there logic every sentence is logically equivalent to a sentence in prenex form. A similar observation regarding first-order formulas under the new stable model semantics of [6] was made in [13]. Thus from a logical point of view there is no loss of generality in defining safety for prenex formulas.

### 3.1 Semi-safety

**Lemma 1.** *If* $\langle (D,\sigma), H, T\rangle \models \varphi$*, then* $\langle (D,\sigma), T, T\rangle \models \varphi$

A variable assignment $\xi$ in a universe $(D,\sigma)$ is a mapping from the set of variables to $D$. If $\varphi \in \mathcal{L}$ has free-variables, $\varphi^{\xi}$ is the closed formula obtained by replacing every free variable $x$ by $\xi(x)$. On the other hand, in the following, if $T \subset At(D,P)$, we denote by $T|_C$ the subset of $T$ whose atoms contain terms only from $\sigma(C)$.

**Lemma 2.** *Let* $\varphi$ *be a quantifier-free formula and* $\xi$ *a variable assignment in a universe* $(D,\sigma)$*. If* $\langle (D,\sigma), T|_C, T\rangle \models \varphi^{\xi}$*, then* $\xi(x) \in \sigma(C)$ *for all* $x \in \mathrm{RV}(\varphi)$*.*

As in [11], we define here a concept of *semi-safety* of a prenex form sentence $\varphi$ in terms of the *semi-safety* of all its variable occurrences. Formally, this is done by defining an operator NSS that collects the variables that have *non-semi-safe* occurrences in a formula $\varphi$.

**Definition 2 (**$NSS$ **and semi-safety).**

1. *If* $\varphi$ *is an atom,* $\mathrm{NSS}(\varphi)$ *is the set of variables in* $\varphi$*.*
2. $\mathrm{NSS}(\varphi_1 \wedge \varphi_2) = \mathrm{NSS}(\varphi_2) \cup \mathrm{NSS}(\varphi_1)$
3. $\mathrm{NSS}(\varphi_1 \vee \varphi_2) = \mathrm{NSS}(\varphi_2) \cup \mathrm{NSS}(\varphi_1)$
4. $\mathrm{NSS}(\varphi_1 \rightarrow \varphi_2) = \mathrm{NSS}(\varphi_2) \setminus \mathrm{RV}(\varphi_1)$.

*The sentence* $\varphi$ *is said to be* semi-safe *if* $\mathrm{NSS}(\varphi) = \varnothing$. □

In other words, a variable $x$ is semi-safe in $\varphi$ if every occurrence is inside some subformula $\alpha \rightarrow \beta$ such that, either $x$ is restricted in $\alpha$ or $x$ is semi-safe in $\beta$. This condition of semi-safety is a relaxation of that of [11], where the implication $\alpha \rightarrow \beta$ should always satisfy that $x$ is restricted in $\alpha$. As a result, (1) is semi-safe, but it is *not* considered so under the definition in [10]. Similarly, our definition implies, for instance, that any occurrence of a variable $x$ in a negated subformula, $\neg\alpha(x)$, will be semi-safe – it corresponds to an implication $\alpha(x) \rightarrow \bot$ with no variables in the consequent. Other examples of semi-safe formulas are, for instance:

$$\neg p(x) \rightarrow (q(x) \rightarrow r(x)) \tag{5}$$

$$p(x) \vee q \rightarrow \neg r(x) \tag{6}$$

Note how in (6), $x$ is not restricted in $p(x) \vee q$ but the consequent $\neg r(x)$ is semi-safe and thus the formula itself. On the contrary, the following formulas are not semi-safe:

$$p(x) \vee q \rightarrow r(x) \tag{7}$$

$$\neg\neg p(x) \wedge \neg r(x) \rightarrow q(x) \tag{8}$$

**Lemma 3.** $\mathrm{NSS}(\varphi_1 \rightarrow \varphi_2) \cup \mathrm{RV}(\varphi_1) \supseteq \mathrm{NSS}(\varphi_2)$.

**Lemma 4.** *Consider a sentence $\varphi$ in prenex form. If $x$ is a variable in $\varphi$ and $x \notin \mathrm{NSS}(\varphi)$, then every occurrence of $x$ is contained in a subformula $\alpha \rightarrow \beta$ such that, if $x$ occurs in $\beta$, then $x \in \mathrm{RV}(\alpha)$.*

Notice that the only rule that forces a variable occurrence to cease being non-semi-safe is $\mathrm{NSS}(\varphi_1 \rightarrow \varphi_2) = \mathrm{NSS}(\varphi_2) \setminus \mathrm{RV}(\varphi_1)$; that is to say, the variable has to appear in the consequent of an implication in whose antecedent the variable is restricted.

**Lemma 5.** *Let $\varphi$ be a quantifier-free formula and $\xi$ a variable assignment such that $\xi(x) \in \sigma(C)$ for all $x \in \mathrm{NSS}(\varphi)$. If $\langle (D,\sigma), T, T\rangle \models \varphi^\xi$, then $\langle (D,\sigma), T|_C, T\rangle \models \varphi^\xi$.*

Proof: By induction over $\psi = \varphi^\xi$. If $\psi$ is atomic the result is trivial and the induction step is also trivial for conjunctive and disjunctive formulas. So, let us assume that $\varphi = \varphi_1 \rightarrow \varphi_2$ where, by the induction hypothesis, $\varphi_2$ is such that, if $\langle (D,\sigma), T, T\rangle \models \varphi_2^\xi$ and $\xi(x) \in \sigma(C)$ for all $x \in \mathrm{NSS}(\varphi_2)$, then $\langle (D,\sigma), T|_C, T\rangle \models \varphi_2^\xi$. Assume

$$\langle (D,\sigma), T, T\rangle \models \varphi_1^\xi \rightarrow \varphi_2^\xi \tag{9}$$

and $\xi(x) \in \sigma(C)$ for all $x \in \mathrm{NSS}(\varphi_1 \rightarrow \varphi_2)$. We must prove that

$$\langle (D,\sigma), T|_C, T\rangle \models \varphi_1^\xi \rightarrow \varphi_2^\xi. \tag{10}$$

First, suppose that $\langle (D,\sigma), T|_C, T\rangle \not\models \varphi_1^\xi$. If $\langle (D,\sigma), T|_C, T\rangle \models \neg\varphi_1^\xi$ then clearly $\langle (D,\sigma), T|_C, T\rangle \models \varphi_1^\xi \rightarrow \varphi_2^\xi$ and we are done. Otherwise, $\langle (D,\sigma), T|_C, T\rangle \models \neg\neg\varphi_1^\xi$. Then, by (9) $\langle (D,\sigma), T|_C, T\rangle, t \models \varphi_1^\xi \rightarrow \varphi_2^\xi$, and since $\langle (D,\sigma), T|_C, T\rangle, h \not\models \varphi_1^\xi$, (10) follows.

Suppose therefore that $\langle (D,\sigma), T|_C, T\rangle \models \varphi_1^\xi$, then $\langle (D,\sigma), T, T\rangle \models \varphi_1^\xi$ and thus

$$\langle (D,\sigma), T, T\rangle \models \varphi_2^\xi \tag{11}$$

On the other hand, if $\langle (D,\sigma), T|_C, T\rangle \models \varphi_1^\xi$, by Lemma 2, $\xi(x) \in \sigma(C)$ for all $x \in \mathrm{RV}(\varphi_1)$ and by hypothesis, $\xi(x) \in \sigma(C)$ for all $x \in \mathrm{NSS}(\varphi_1 \rightarrow \varphi_2)$; in particular, for $\mathrm{NSS}(\varphi_1 \rightarrow \varphi_2) \cup \mathrm{RV}(\varphi_1) \supseteq \mathrm{NSS}(\varphi_2)$ (lemma 3), we have:

$$\xi(x) \in \sigma(C) \text{ for all } x \in \mathrm{NSS}(\varphi_2) \tag{12}$$

By the induction hypothesis, (11) and (12), we conclude that $\langle (D,\sigma), T|_C, T\rangle \models \varphi_2^\xi$. □

**Lemma 6.** *If $\varphi$ is semi-safe, and $\langle (D,\sigma), T, T\rangle \models \varphi$, then $\langle (D,\sigma), T|_C, T\rangle \models \varphi$.*

Proof: We can assume that $\varphi$ is a prenex form and then, the proof is trivial by induction over the length of its prefix.

If $n = 0$, the formula is quantifier-free and we are in the situation of the previous lemma with $\mathrm{NSS}(\varphi) = \varnothing$. Let us assume that the result is valid for prenex formulas whose prefix has length $n-1$ and let $\varphi = Q_n x_n \dots Q_1 x_1 \varphi(x_1, \dots, x_n)$.

- $\varphi = \exists x \psi(x)$: $\langle (D,\sigma), T, T\rangle \models \exists x \psi(x)$ iff $\langle (D,\sigma), T, T\rangle \models \psi(d)$ for some $d \in D$. Then, $\langle (D,\sigma), T|_C, T\rangle \models \psi(d)$ and thus $\langle (D,\sigma), T|_C, T\rangle \models \exists x \psi(x)$
- $\varphi = \forall x \psi(x)$: $\langle (D,\sigma), T, T\rangle \models \forall x \psi(x)$ iff $\langle (D,\sigma), T, T\rangle \models \psi(d)$ for every $d \in D$. Then, $\langle (D,\sigma), T|_C, T\rangle \models \psi(d)$ for every $d \in D$ and thus $\langle (D,\sigma), T|_C, T\rangle \models \forall x \psi(x)$. $\square$

**Theorem 2.** *If $\varphi$ is semi-safe, and $\langle (D,\sigma), T, T\rangle$ is an equilibrium model of $\varphi$, then $T|_C = T$.*

### 3.2 safe formulas

The concept of safety relies on semi-safety plus an additional condition on variable occurrences that can be defined in terms of Kleene's three-valued logic [**?**]. Given a three-valued interpretation $\nu \colon At \to \{0, 1/2, 1\}$, we extend it to evaluate arbitrary formulas $\nu(\varphi)$ as follows:

$$\begin{array}{ll} \nu(\varphi \wedge \psi) = \min(\nu(\varphi), \nu(\psi)) & \nu(\bot) = 0 \\ \nu(\varphi \vee \psi) = \max(\nu(\varphi), \nu(\psi)) & \nu(\varphi \to \psi) = \max(1 - \nu(\varphi), \nu(\psi)) \end{array}$$

from which we can derive $\nu(\neg\varphi) = \nu(\varphi \to \bot) = 1 - \nu(\varphi)$ and $\nu(\top) = \nu(\neg\bot) = 1$.

**Definition 3 ($\nu_x$ operator).** *Given any quantifier-free formula $\varphi$ and any variable $x$, we define the three-valued interpretation so that for any atom $\alpha$, $\nu_x(\alpha) = 0$ if $x$ occurs in $\alpha$ and $\nu_x(\alpha) = 1/2$ otherwise.*

Intuitively, $\nu_x(\varphi)$ fixes all atoms containing the variable $x$ to 0 (falsity) leaving all the rest undefined and then evaluates $\varphi$ using Kleene's three-valued operators, that is nothing else but exploiting the defined values 1 (true) and 0 (false) as much as possible. For instance, $\nu_x(p(x) \to q(x))$ would informally correspond to $\nu_x(0 \to 0) = \max(1 - 0, 0) = 1$ whereas $\nu_x(p(x) \vee r(y) \to q(x)) = \nu_x(0 \vee 1/2 \to 0) = \max(1 - \max(0, 1/2), 0) = 1/2$.

For the following definition, we need to recall the following terminology: a subexpression of a formula is said to be positive in it if the number of implications that contain that subexpression in the antecedent is even, and negative if it is odd.

**Definition 4 (Weakly-restricted variable).** *An occurrence of a variable $x$ in $Qx\ \varphi$ is* weakly-restricted *if it occurs in a subformula $\psi$ of $\varphi$ such that:*

- *$Q = \forall$, $\psi$ is positive and $\nu_x(\psi) = 1$*
- *$Q = \forall$, $\psi$ is negative and $\nu_x(\psi) = 0$*
- *$Q = \exists$, $\psi$ is positive and $\nu_x(\psi) = 0$*
- *$Q = \exists$, $\psi$ is negative and $\nu_x(\psi) = 1$*

*In all cases, we further say that $\psi$* makes the ocurrence weakly restricted in $\varphi$.

For the following lemmas we use the many-valued semantics for here-and-there logic with $\{0, 1, 2\}$ as truth values; the order relation $\leq$ is the usual one in this set [17]. The **QHT**-interpretations are translated to many-valued assignments with interesting properties wrt the order $\leq$; for example, the interpretations of $\wedge$ and $\vee$ are increasing in both arguments, $\neg$ is decreasing and $\to$ is incresing in second argument and decreasing in the antecedent. The next lemma is a consequence of these properties and will be used later.

**Lemma 7.** *If $\mathcal{M}$ is a many valued assignment corresponding to the structure $\langle (D,\sigma), H, T\rangle$ then $\mathcal{M}$ is increasing in the positive subformulas and decreasing in the negative subformulas.*

The Kleene semantics and the three-valued semantics of here-and-there agree for top and bottom values and thus we can conclude the following result.

**Lemma 8.** *Let $\psi(x)$ be a quantifier-free formula and let $\langle (D,\sigma), H, T\rangle$ be such that $T \subset \mathrm{At}(\sigma(C), P)$. Then:*

- *If $d \notin \sigma(C)$ and $\nu_x(\psi) = 1$, then $\langle (D,\sigma), H, T\rangle \models \psi(d)$*
- *If $d \notin \sigma(C)$ and $\nu_x(\psi) = 0$, then $\langle (D,\sigma), H, T\rangle \models \neg\psi(d)$*

**Definition 5.** *A semi-safe sentence is said to be* safe *if all its positive occurrences of universally quantified variables, and all its negative occurrences of existentially quantified variables are weakly restricted.*

For instance, the formula $\varphi = \forall x(\neg q(x) \rightarrow (r \vee \neg p(x)))$ is safe: the occurrence of $x$ in $p(x)$ is negative, whereas the occurrence in $q(x)$ is inside a positive subformula, $\varphi$ itself, for which $x$ is weakly-restricted, since $\nu_x(\varphi) = \neg 0 \rightarrow (1/2 \vee \neg 0) = 1$. The occurrence of $x$ in $\neg q(x)$ is not restricted and as a consequence the formula is not LLP-safe. The other aspect where Definition 5 is more general than LLP-safety results from the fact that to be safe $x$ can occur negatively in $\varphi$ given $Q_i = \forall$ or positively given $Q_i = \exists$. Another example of a safe formula is $\forall x((\neg\neg p(x) \wedge q(x)) \rightarrow r)$.

**Lemma 9.** *Let $\varphi(x)$ a prenex formula that has no free variables other than $x$. Let $\langle (D,\sigma), H, T\rangle$ be a model such that $T \subset \mathrm{At}(\sigma(C), P)$. Then:*

1. *If $\forall x\varphi(x)$ is safe:* $\langle (D,\sigma), H, T\rangle \models \forall x\varphi(x)$ *iff* $\langle (D,\sigma), H, T\rangle \models \bigwedge\limits_{c\in C} \varphi(c)$.
2. *If $\exists x\varphi(x)$ is safe:* $\langle (D,\sigma), H, T\rangle \models \exists x\varphi(x)$ *iff* $\langle (D,\sigma), H, T\rangle \models \bigvee\limits_{c\in C} \varphi(c)$.

*Proof.* The proof makes use of the monotonic properties of the many-valued assignments with respect to the positive and negative ocurrences of subformulas.

1. For item 1. we only need to prove the sufficient condition. Let us assume that $\langle (D,\sigma), H, T\rangle \models \bigwedge_{c\in C} \varphi(c)$ and let $d \in D$. If $d \in \sigma(C)$, then $\langle (D,\sigma), H, T\rangle \models \varphi(d)$. If $d \notin \sigma(C)$, we consider the maximal subformulas $\psi_i$, $i = 1, \ldots, n$, and $\gamma_i$, $i = 1, \ldots, m$, making weakly restricted the positive occurrence $x$ in $\varphi(x)$ and such that every $\psi_i$ is positive and every $\gamma_i$ is negative en $\varphi$; the other occurrences of $x$ are included in atoms, $\alpha_i(x)$, $i = 1, \ldots, k$, which occur negatively in $\varphi(x)$. Let $\mathcal{M}$ be the structure $\langle (D,\sigma), H, T\rangle$ as a many-valued assignment and $c \in \sigma(C)$. By lemma 8: for every $\psi_i$, $\langle (D,\sigma), H, T\rangle \models \psi_i(d)$ and thus $\mathcal{M}(\psi_i(d)) = 2 \geq \mathcal{M}(\psi_i(c))$; for every $\gamma_i$, $\langle (D,\sigma), H, T\rangle \models \neg\gamma_i(d)$ and thus $\mathcal{M}(\gamma_i(d)) = 0 \leq \mathcal{M}(\gamma_i(c))$; on the other hand, $\mathcal{M}(\alpha_i(d)) \leq \mathcal{M}(\alpha_i(c))$. Therefore, by lemma 7, $\mathcal{M}(\varphi(d)) \geq \mathcal{M}(\varphi(c)) = 2$, because every occurrence of $x$ is include either in some $\psi_i$ or some $\gamma_i$ or some $\alpha_i$. That is, we have proved that $\langle (D,\sigma), H, T\rangle \models \varphi(d)$ for every $d \in D$ and thus $\langle (D,\sigma), H, T\rangle \models \forall x\varphi(x)$.
2. The proof of item 2. is similar. □

## 4 Grounding

Let $(D,\sigma)$ be a finite domain and $D' \subseteq D$; the grounding over $D'$ of a sentence $\varphi$ is defined recursively by:

$$\begin{aligned}
\mathrm{Gr}_{D'}(\varphi) &= \varphi \text{ if } \varphi \text{ is quantifier-free}\\
\mathrm{Gr}_{D'}(\varphi_1 \odot \varphi_2) &= \mathrm{Gr}_{D'}(\varphi_1) \odot \mathrm{Gr}_{D'}(\varphi_2) \quad \odot \in \{\rightarrow, \wedge, \vee\}\\
\mathrm{Gr}_{D'}(\forall x\varphi(x)) &= \bigwedge_{d\in D'} \mathrm{Gr}_{D'}\varphi(d)\\
\mathrm{Gr}_{D'}(\exists x\varphi(x)) &= \bigvee_{d\in D'} \mathrm{Gr}_{D'}\varphi(d)
\end{aligned}$$

The following property of grounding can be shown.[2]

**Proposition 1.** $\langle (D,\sigma), H, T\rangle \models \varphi$ *if and only if* $\langle (D,\sigma), H, T\rangle \models \mathrm{Gr}_D(\varphi)$

[2] A version of this property was shown for universal theories in [17].

In particular, if we work with PNA-models ($D = \sigma(C)$), we can do the grounding using the initial constants. But we must work with the first order semantics, because with the mapping $\sigma$ two constants may denote the same element of the domain; then we would obtain:

$$\langle (D,\sigma), H, T\rangle \models \varphi \text{ iff } \langle (D,\sigma), H, T\rangle \models \mathrm{Gr}_C(\varphi) \tag{13}$$

**Lemma 10.** *If $T \subset At(\sigma(C), P)$, $d \notin \sigma(C)$, and $\langle (D,\sigma), H, T\rangle \models \varphi(d)$, it follows that $\langle (D,\sigma), H, T\rangle \models \varphi(d')$ for every $d' \in D \setminus \sigma(C)$.*

This result is trivial, because the meaning of the atoms in $\varphi(d)$ and $\varphi(d')$ in the interpretation is the same, for $T$ does not contain elements outside $\sigma(C)$. Thus, if we are interested only in Herbrand models, we can adopt a reduced grounding, even if we consider general domains.

**Lemma 11.** *If $T \subset At(\sigma(C), P)$ and there exists $a \in D \setminus \sigma(C)$ then,*

1. *$\langle (D,\sigma), H, T\rangle \models \forall x\varphi(x)$ if and only if $\langle (D,\sigma), H, T\rangle \models \bigwedge_{d\in\sigma(C)\cup\{a\}} \varphi(d)$*
2. *$\langle (D,\sigma), H, T\rangle \models \exists x\varphi(x)$ if and only if $\langle (D,\sigma), H, T\rangle \models \bigvee_{d\in\sigma(C)\cup\{a\}} \varphi(d)$*

This result is trivial from the previous lemma and allows us to conclude the following generalisation of property (13).

**Proposition 2.** *If $T \subset At(\sigma(C), P)$ and $a \in D \setminus \sigma(C)$, then $\langle (\sigma, D), H, T\rangle \models \varphi$ if and only if $\langle (\sigma, D), H, T\rangle \models \mathrm{Gr}_{C\cup\{a\}}(\varphi)$.*

**Theorem 3.** *Let $\varphi$ be a safe prenex formula. Let $\langle (D,\sigma), H, T\rangle$ be a model such that $T \subset \mathrm{At}(\sigma(C), P)$. Then:*

$\langle (D,\sigma), H, T\rangle \models \varphi$ *if and only if* $\langle (D,\sigma), H, T\rangle \models \mathrm{Gr}_C(\varphi)$

*Proof.* By induction on the length of the prefix.

We can now show as promised the property that safe formulas satisfy domain independence. The following is immediate from Theorems 2 and 3 and the definition of equilibrium model.

**Proposition 3.** *Let $\varphi$ be a safe prenex formula, then: $\langle (D,\sigma), T, T\rangle$ is an equilibrium model of $\varphi$ if and only if it is an equilibrium model of $\mathrm{Gr}_C(\varphi)$.*

Here is an alternative formulation using language expansions.

**Theorem 4.** *Let $\varphi$ be a safe prenex formula. Suppose we expand the language $\mathcal{L}$ by considering a set of constants $C' \supset C$. A total* **QHT***-model $\langle (D,\sigma), T, T\rangle$ is an equilibrium model of $\mathrm{Gr}_{C'}(\varphi)$ if an only if it is an equilibrium model of $\mathrm{Gr}_C(\varphi)$.*

*Proof.* Consider the grounding of the safe formula $\varphi$ wrt the expanded language. Suppose that $\mathcal{M} = \langle (D,\sigma), T, T\rangle$ is a total model of $\mathrm{Gr}_{C'}(\varphi)$. Since $C \subset C'$ we have $\sigma(C) \subseteq \sigma(C') \subseteq D$. By Theorem 2 and an obvious extension of Proposition 2, $\mathcal{M}$ is an equilibrium model of $\varphi$ iff it is an equilibrium model of $\mathrm{Gr}_{C'}(\varphi)$. And by Proposition 3, $\mathcal{M}$ is an equilibrium model of $\varphi$ iff it is an equilibrium model of $\mathrm{Gr}_C(\varphi)$. From this it follows that the equilibrium models of $\mathrm{Gr}_C(\varphi)$ and $\mathrm{Gr}_{C'}(\varphi)$ coincide.

The concept of domain independence from [2] is obtained as a special case if we assume that the unique name assumption applies.

## 5 Definability

An important property of safe formulas, as we shall now show, is that their equilibrium models form a definable class in first-order logic;[3] more specifically that we can effectively exhibit a ground, first-order formula whose models are precisely the equilibrium models of our theory. An alternative approach to exhibiting definable classes of stable models, using loop formulas, can be found in [12].

Let $\varphi$ be a ground sentence and $(\alpha_1, \ldots, \alpha_n)$ the sequence of atoms in $\varphi$. If $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_n)$ is another sequence of ground atoms, the formula $\varphi[\boldsymbol{\beta}]$ is built recursively:

- $\alpha_i[\boldsymbol{\beta}] = \beta_i$
- $\bot[\boldsymbol{\beta}] = \bot$
- $(\psi_1 \wedge \psi_2)[\boldsymbol{\beta}] = \psi_1[\boldsymbol{\beta}] \wedge \psi_2[\boldsymbol{\beta}]$
- $(\psi_1 \vee \psi_2)[\boldsymbol{\beta}] = \psi_1[\boldsymbol{\beta}] \vee \psi_2[\boldsymbol{\beta}]$
- $(\psi_1 \rightarrow \psi_2)[\boldsymbol{\beta}] = (\psi_1[\boldsymbol{\beta}] \rightarrow \psi_2[\boldsymbol{\beta}]) \wedge (\varphi \rightarrow \psi)$
- $(\neg\psi)[\boldsymbol{\beta}] = \neg\psi[\boldsymbol{\beta}] \wedge \neg\psi$

**Lemma 12.** *Let $\varphi$ be a ground sentence and $(\alpha_1, \ldots, \alpha_n)$ the sequence of atoms in $\varphi$. Let $U$ be the sequence $(u_1, \ldots, u_n)$ such that, for every $i$, either $u_i = \bot$ or $u_i = \alpha_i$ and there is some $i$ such that $u_i = \bot$. $\langle (D, \sigma), T, T\rangle$ is an equilibrium model of $\varphi$ if and only if $T$ is a classical model of $\varphi \wedge \neg \bigvee_{\mathbf{u} \in U} \varphi[\mathbf{u}]$.*

Proof: Let $\langle (D, \sigma), T, T\rangle$ be a model of $\varphi$; for every $\mathbf{u} \in U$, we define

$$H_{\mathbf{u}} = \{\ell \in T \mid \text{ exists } i, \text{ such that } \sigma(u_i) = \ell\}$$

where $\sigma(p(t_1, \ldots, t_m)) = p(\sigma(t_1), \ldots, \sigma(t_m))$. Then $H_{\mathbf{u}} \subset T$ and $\langle (D, \sigma), H_{\mathbf{u}}, T\rangle \not\models \varphi$. So we have defined a bijection between the $U$ and the set of $\mathbf{QHT}$-interpretations strictly less than $\langle (D, \sigma), T, T\rangle$. This bijection verifies the following property:

$$\langle (D, \sigma), H_{\mathbf{u}}, T\rangle \models \varphi \Leftrightarrow T \models \varphi[\mathbf{u}] \tag{14}$$

The relation (14) can be easily proved by induction over $\varphi$ and allows us to conclude that, if $\langle (D, \sigma), H_{\mathbf{u}}, T\rangle$ is an equilibrium model of $\varphi$ then every formula $\varphi[\mathbf{u}]$ is not valid in $T$ and thus $T \models \varphi \wedge \neg \bigvee_{\mathbf{u} \in U} \varphi[\mathbf{u}]$. $\square$

Satisfaction in classical logic can be encoded by satisfaction by total models in here-and-there logic [19]; so we can characterise the classical models of the formula in the previous lemma by the total models of a formula in $\mathbf{QHT}$, as follows:

**Corollary 1.** *Let $\varphi$ be a ground sentence and $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_n)$ the sequence of atoms in $\varphi$. Let $U$ be the sequence $(u_1, \ldots, u_n)$ such that, for every $i$, either $u_i = \bot$ or $u_i = \alpha_i$ and there is some $i$ such that $u_i = \bot$. $\langle (D, \sigma), T, T\rangle$ is an equilibrium model of $\varphi$ if and only if it is a $\mathbf{QHT}$-model of*

$$\bigwedge_{1 \leq i \leq n} (\neg\neg\alpha_i \rightarrow \alpha_i) \wedge \varphi[\neg\neg\boldsymbol{\alpha}] \wedge \neg \bigvee_{\mathbf{u} \in U} \varphi[\mathbf{u}]$$

As a consequence of this and Theorem 3 we obtain:

**Theorem 5.** *For every safe sentence $\varphi$, there exists a ground formula $\psi$ in the same language such that $\langle (D, \sigma), T, T\rangle$ is an equilibrium model of $\varphi$ if and only if it is a $\mathbf{QHT}$-model of $\psi$.*

[3] Shown for LLP-safety in [**?**]

## 6 Discussion and relation to other concepts of safety

As remarked in the introduction, one of our aims is to generalise the safety concept of [10,11] and at the same time capture logic programs that are safe according to [2]. Although for reasons of space we cannot present here the full definition of LLP-safety from [11], we explained after Definition 5 the main features of our concept that generalise that of [11] (see also the examples discussed below). To verify the second objective we need to consider the class of logic programs treated by Bria, Faber and Leone in [2]. These programs are called *normal form nested*, or NFN for short. They comprise rules of a special form; in logical notation they are formulas that have the shape:

$$\bigwedge_{1\leq i\leq m} \Delta_i \rightarrow \bigvee_{\leq i\leq n} \Gamma_j \tag{15}$$

where each $\Delta_i$ is a disjunction of literals and each $\Gamma_j$ is a conjunction of atoms. A $\Delta_i$ is called a *basic* disjunction and is said to be *positive* if it contains only positive literals or atoms. In a rule $r$ of form (15) the antecedent is called the *body* and the consequent the *head* of the rule. According to [2] $r$ is safe if each variable in its head literals and its negative body literals is safe, where a variable $x$ is said to be safe in $r$ if if there exists a positive basic disjunction $\Delta_i$ in the body of $r$ such that $x$ is a variable of $\alpha$ for every atom $\alpha \in \Delta_i$.

**Proposition 4.** *All rules of an NFN program that are safe according to [2] are safe formulas in the sense of Definition 5.*

*Proof.* Each rule of form (15) can be seen as a formula in prenex form all of whose quantifiers are universal. By BFL-safety, each variable of $r$ is contained in a positive subformula $\alpha \rightarrow \beta$ which is $r$ itself. Moreover $x$ occurs in every atom of some positive basic disjunction $\Delta_i$, so $x \in \mathrm{RV}(\Delta_i)$. It follows that $x$ is weakly restricted in the rule. □

Let us consider some examples showing that the safety concept introduced in Definition 5 is quite general and, in many cases, is better behaved with respect to strongly equivalent transformations than other definitions presented before.

There exist more examples in which our definition is capable of detecting safe sentences that LLP-safety cannot. For instance, although

$$\neg\neg p(x) \rightarrow q \tag{16}$$

is LLP-safe, once we add $\neg r(x)$ to the body

$$\neg\neg p(x) \wedge \neg r(x) \rightarrow q \tag{17}$$

it becomes an LLP-unsafe sentence, whereas it still preserves safety under the current definition. Notice that the new $x$ occurs negatively in the body but it is "made safe" by $\neg\neg p(x)$ and the fact that $x$ does not occur in the head. In fact, if the head becomes $q(x)$ like in (17), the formula is not domain independent and so is not LLP-safe nor safe under our definition (in fact, as we saw, it is not even semi-safe).

The idea of modifying the definition of semi-safe formulas to check, as in [2], that a variable is restricted only if occurs in the head, can also be applied to existentially quantified formulas. For instance, this formula

$$\exists x\ (\neg\neg(p(x) \vee q)) \tag{18}$$

should be safe. It is strongly equivalent to $\neg\exists x\ p(x) \rightarrow q$ and this can be easily captured by introducing an auxiliary predicate in a safe way as follows:

$$\begin{aligned} aux_1 &\leftarrow p(X) \\ q &\leftarrow \neg aux_1 \end{aligned}$$

In fact, (18) is safe. However, it is not even semi-safe according to the definition in [10]: there is no implication containing $x$ restricted in the antecedent.

Finally, as an example in which LLP-safety varies under a strongly equivalent transformation, take the formula (5) which is safe under our concept, but not LLP-safe. This kind of nested implication in the head can be simply reduced, under strong equivalence, to $\neg p(x) \wedge q(x) \rightarrow r(x)$ which is still safe under our definition and becomes LLP-safe too.

## 7 Conclusion

We have presented a new concept of safety for general first-order formulas under stable model semantics and hence for very general kinds of answer set programs. As a logical formalism we have used quantified equilibrium logic based on the non-classical logic of quantified here-and-there, **QHT**, in which equilibrium or stable models can be easily defined as minimal models and their logical properties readily studied. We showed that the new concept of safety extends and generalises recent definitions of safety from [11] and [2]. Unlike in [2], in our approach we do not have to make the unique name assumption. We showed that safe formulas satisfy three main properties or criteria of adequacy: their stable models do not contain unnamed individuals, they satisfy the condition of domain independence, and they form a first-order definable class of structures.

One topic for future work concerns program transformations. While it is clear that many strong-equivalence preserving transformations (eg. from [4,5]) preserve safety under our definition, we do not have yet a full picture of precisely which syntactic classes have this property.

## References


1. Baral, C.: Knowledge Representation, Reasoning and Declarative Problem Solving. CUP, Cambridge (2002)
2. A. Bria, W. Faber and N. Leone. Normal Form Nested Programs. S. Hölldobler *et al* (eds), JELIA 2008, Springer LNAI 5293, pp. 76–88, 2008.
3. de Bruijn, J., Pearce, D., Polleres, A., Valverde, A.: Quantified Equilibrium Logic and Hybrid Rules. In: Marchiori, M. *et al* (eds), RR2007, LNCS vol 4524, pp. 58-72, Springer, Heidelberg (2007)
4. Cabalar, P., Pearce, D., Valverde, A.: Reducing Propositional Theories in Equilibrium Logic to Logic Programs. In: Bento, C. *et al* (eds) EPIA 05, LNAI vol 3808, pp. 4-17 Springer, Heidelberg (2005)
5. Cabalar, P., Pearce, D., Valverde, A.: Minimal Logic Programs. In: Dahl, V., Niemelä, I. (eds) ICLP 07, LNCS 4670, pp. 104-118, Springer, Heidelberg (2007).
   Proceedings of the Twentieth National Conference on Artificial Intelligence, AAAI-2005, pp. 695-700
6. Ferraris, P., Lee, J., Lifschitz, V.: A New Perspective on Stable Models. In: Veloso, M. (ed) 20th International Joint Conference on Artificial Intelligence, IJCAI 2007, pp. 372-379.
7. Heymans, S., van Nieuwenborgh, D., Vermeir, D.: Open Answer Set Programming with Guarded Programs. ACM Trans. Comp. Logic 9, Article 26 (2008)
8. Stijn Heymans, Davy Van Nieuwenborgh and Dirk Vermeir. Guarded Open Answer Set Programming. In *Proc. of 8th International Conference on Logic Programming and Non Monotonic Reasoning (LPNMR 2005)*, Springer LNAI 3662, 2005.
9. Heymans, S., Predoiu, L., Feier, C., de Bruijn, J., van Nieuwenborgh, D.: G-hybrid Knowledge Bases. In: Applications of Logic Programming in the Semantic Web and Semantic Web Services, ALPSWS 2006, CEUR Workshop Proceedings, vol 196 (2006)
10. Lee, J., Lifschitz, V., Palla, R.: A Reductive Semantics for Counting and Choice in Answer Set Programming. In: Proceedings of the 23rd AAAI Conference on Artificial Intelligence, AAAI 2008, pp. 472-479.
11. J. Lee, V. Lifschitz and R. Palla. Safe formulas in the general theory of stable models. (Preliminary report). in Proceedings of ICLP-08, Springer, LNCS, 2008.
12. J. Lee and V. Meng. On Loop Formulas with Variables. Proceedings of AAAI-08, AAAI. 2008.
13. Lee, J. Palla, R. Yet Another Proof of the Strong Equivalence between Propositional Theories and Logic Programs. In: Correspondence and Equivalence between Nonmonotonic Theories, CENT 2007, volume 265 of CEUR Workshop Proceedings, Tempe, AZ, May 2007. CEUR-WS.org.
14. V. Lifschitz, D. Pearce, and A. Valverde. A Characterization of Strong Equivalence for Logic Programs with Variables. In *Proc. of LPNMR 2007*, Springer LNAI 4483, pp. 188-200.
15. V. Lifschitz, L. Tang and H. Turner. Nested Expressions in Logic Programs. *Annals of Mathematics and Artificial Intelligence*, 25(3-4): 369-389, 1999.
16. D. Pearce, A. Valverde. Towards a first order equilibrium logic for nonmonotonic reasoning. In *Proc. of JELIA 2004*, Springer LNAI 3229, pp. 147-160.
17. D. Pearce, A. Valverde. A First-Order Nonmonotonic Extension of Constructive Logic. *Studia Logica* 80:321-346, 2005.
18. D. Pearce, A. Valverde. Quantified Equilibrium Logic. *Tech. report, Univ. Rey Juan Carlos*, 2006. $http : //www.matap.uma.es/investigacion/tr/ma06_02.pdf$.
19. D. Pearce, A. Valverde. Quantified Equilibrium Logic and Foundations for Answer Set Programs. *Proc. ICLP 08*, Springer, LNCS, 2008.
20. D. van Dalen. *Logic and Structure*. Springer, 1983.