



HAL
open science

Founded World Views with Autoepistemic Equilibrium Logic

Pedro Calabar, Jorge Fandinno, Luis Fariñas del Cerro

► **To cite this version:**

Pedro Calabar, Jorge Fandinno, Luis Fariñas del Cerro. Founded World Views with Autoepistemic Equilibrium Logic. International Conference on Logic Programming and Nonmonotonic Reasoning (LPNMR 2019), Jun 2019, Philadelphia, PA, United States. pp.134-147, 10.1007/978-3-030-20528-7_11 . hal-02378389

HAL Id: hal-02378389

<https://hal.science/hal-02378389v1>

Submitted on 25 Nov 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Open Archive Toulouse Archive Ouverte

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible

This is an author's version published in:

<http://oatao.univ-toulouse.fr/24816>

Official URL

DOI : https://doi.org/10.1007/978-3-030-20528-7_11

To cite this version: Calabar, Pedro and Fandinno, Jorge and Fariñas del Cerro, Luis *Founded World Views with Autoepistemic Equilibrium Logic*. (2019) In: Logic Programming and Nonmonotonic Reasoning (LPNMR 2019), 3 June 2019 - 7 June 2019 (Philadelphia, PA, United States).

Any correspondence concerning this service should be sent to the repository administrator: tech-oatao@listes-diff.inp-toulouse.fr

Founded World Views with Autoepistemic Equilibrium Logic

Pedro Cabalar¹(✉), Jorge Fandinno², and Fariñas del Cerro Luis²

¹ University of A Coruña, A Coruña, Spain
cabalar@udc.es

² IRIT University of Toulouse, CNRS, Toulouse, France
{jorge.fandinno, farinas}@irit.fr

Abstract. Defined by Gelfond in 1991 (G91), epistemic specifications (or programs) are an extension of logic programming under stable models semantics that introduces *subjective literals*. A subjective literal allows checking whether some regular literal is true in all (or in some of) the stable models of the program, being those models collected in a set called *world view*. One epistemic program may yield several world views but, under the original G91 semantics, some of them resulted from self-supported derivations. During the last eight years, several alternative approaches have been proposed to get rid of these self-supported world views. Unfortunately, their success could only be measured by studying their behaviour on a set of common examples in the literature, since no formal property of “self-supportedness” had been defined. To fill this gap, we extend in this paper the idea of unfounded set from standard logic programming to the epistemic case. We define when a world view is *founded* with respect to some program and propose the *foundedness* property for any semantics whose world views are always founded. Using counterexamples, we explain that the previous approaches violate foundedness, and proceed to propose a new semantics based on a combination of Moore’s Autoepistemic Logic and Pearce’s Equilibrium Logic. The main result proves that this new semantics precisely captures the set of founded G91 world views.

1 Introduction

The language of *epistemic specifications*, proposed by Gelfond in 1991 [4], extends disjunctive logic programs (under the *stable model* [6] semantics) with modal constructs called *subjective literals*. Using these constructs, it is possible to check

This work was partially supported by MINECO, Spain, grant TIC2017-84453-P, Xunta de Galicia, Spain (GPC ED431B 2016/035 and 2016–2019 ED431G/01, CITIC). The second author is funded by the Centre International de Mathématiques et d’Informatique de Toulouse (CIMI) through contract ANR-11-LABEX-0040-CIMI within the program ANR-11-IDEX-0002-02.

whether a regular literal l is true in *every* stable model (written $\mathbf{K}l$) or in *some* stable model (written $\mathbf{M}l$) of the program. For instance, the rule:

$$a \leftarrow \neg \mathbf{K} b \tag{1}$$

means that a must hold if we cannot prove that all the stable models contain b . The definition of a “satisfactory” semantics for epistemic specifications has proved to be a non-trivial enterprise, as shown by the list of different attempts proposed so far [2,4,5,9,20–22]. The main difficulty arises because subjective literals query the set of stable models but, at the same time, occur in rules that determine those stable models. As an example, the program consisting of:

$$b \leftarrow \neg \mathbf{K} a \tag{2}$$

and (1) has now two rules defining atoms a and b in terms of the presence of those same atoms in all the stable models. To solve this kind of cyclic interdependence, the original semantics by Gelfond [4] (G91) considered different alternative *world views* or sets of stable models. In the case of program (1)–(2), G91 yields two alternative world views¹, $[\{a\}]$ and $[\{b\}]$, each one containing a single stable model, and this is also the behaviour obtained in the remaining approaches developed later on. The feature that made G91 unconvincing, though, was the generation of self-supported world views. A prototypical example for this effect is the epistemic program consisting of the single rule:

$$a \leftarrow \mathbf{K} a \tag{3}$$

whose world views under G91 are $[\emptyset]$ and $[\{a\}]$. The latter is considered as counter-intuitive by all authors² because it relies on a self-supported derivation: a is derived from $\mathbf{K} a$ by rule (3), but the only way to obtain $\mathbf{K} a$ is rule (3) itself. Although the rejection of world views of this kind seems natural, the truth is that all approaches in the literature have concentrated on studying the effects on individual examples, rather than capturing the absence of self-supportedness as a formal property. To achieve such a goal, we would need to establish some kind of *derivability* condition in a very similar fashion as done with *unfounded sets* [3] for standard logic programs. To understand the similarity, think about the (tautological) rule $a \leftarrow a$. The classical models of this rule are \emptyset and $\{a\}$, but the latter cannot be a stable model because a is not derivable applying the rule. Intuitively, an unfounded set is a collection of atoms that is not derivable from a given program and a fixed set of assumptions, as happens to $\{a\}$ in the last example. As proved by [12], the stable models of any disjunctive logic program are precisely its classical models that are *founded*, that is, that do not admit any unfounded set. As we can see, the situation in (3) is pretty similar to $a \leftarrow a$ but, this time, involves derivability through subjective literals. An immediate option

¹ For the sake of readability, sets of propositional interpretations are embraced with $[]$ rather than $\{ \}$.

² This includes Gelfond himself, who proposed a new variant in [5] motivated by this same example and further modified this variant later on in [9].

is, therefore, extending the definition of unfounded sets for the case of epistemic programs – this constitutes, indeed, the first contribution of this paper.

Once the property of *founded* world views is explicitly stated, the paper proposes a new semantics for epistemic specifications, called *Founded Autoepistemic Equilibrium Logic* (FAEEL), that fulfills that requirement. In the spirit of [2, 22], our proposal actually constitutes a full modal non-monotonic logic where \mathbf{K} becomes the usual necessity operator applicable to arbitrary formulas. Formally, FAEEL is a combination of Pearce’s *Equilibrium Logic* [17], a well-known logical characterisation of stable models, with Moore’s *Autoepistemic Logic* (AEL) [15], one of the most representative approaches among modal non-monotonic logics. The reason for choosing Equilibrium Logic is quite obvious, as it has proved its utility for characterising other extensions of ASP, including the already mentioned epistemic approaches in [2, 22]. As for the choice of AEL, it shares with epistemic specifications the common idea of *agent’s introspection* where $\mathbf{K}\varphi$ means that φ is one of the agent’s beliefs. The only difference is that those beliefs are just classical models in the case of AEL whereas epistemic specifications deal with stable models instead. Interestingly, the problem of self-supported models has also been extensively studied in AEL [11, 13, 16, 18], where the formula $\mathbf{K}a \rightarrow a$, analogous to (3), also yields an unfounded world view³ $\{\{a\}\}$. Our solution consists in combining the monotonic bases of AEL and Equilibrium Logic (the modal logic KD45 and the intermediate logic of Here-and-There (HT) [8], respectively), but defining a two-step models selection criterion that simultaneously keeps the agent’s beliefs as stable models and avoids unfounded world views from the use of the modal operator \mathbf{K} . As expected, we prove that FAEEL guarantees the property of founded world views, among other features. Our main result, however, goes further and asserts that the FAEEL world views of an epistemic program are precisely the set of founded G91 world views. We reach, in this way, an analogous situation to the case of standard logic programming, where stable models are the set of founded classical models of the program.

The rest of the paper is organised as follows. Section 2 and 3 respectively revisit the background knowledge about equilibrium logic and epistemic specifications necessary for the rest of the paper. Section 4 introduces the foundedness property for epistemic logic programs. In Sect. 5, we introduce FAEEL and show that its world views precisely coincide with the set of founded G91 world views. Finally, Sect. 6 concludes the paper.

2 Background

We begin recalling the basic definitions of equilibrium logic and its relation to stable models. We start from the syntax of propositional logic, with formulas built from combinations of atoms in a set \mathcal{AT} with operators \wedge, \vee, \perp and \rightarrow in the usual way. We define the derived operators $\varphi \leftrightarrow \psi \stackrel{\text{def}}{=} (\varphi \rightarrow \psi) \wedge (\psi \rightarrow \varphi)$, $(\varphi \leftarrow \psi) \stackrel{\text{def}}{=} (\psi \rightarrow \varphi)$, $\neg\varphi \stackrel{\text{def}}{=} (\varphi \rightarrow \perp)$ and $\top \stackrel{\text{def}}{=} \neg\perp$.

³ Technically, AEL is defined in terms of *theory expansions* but each one can be characterised by a canonical S5-model with the same form of a world view [14, 19].

A *propositional interpretation* T is a set of atoms $T \subseteq \mathcal{AT}$. We write $T \models \varphi$ to represent that T classically satisfies formula φ . An *HT-interpretation* is a pair $\langle H, T \rangle$ (respectively called “here” and “there”) of propositional interpretations such that $H \subseteq T \subseteq \mathcal{AT}$; it is said to be *total* when $H = T$. We write $\langle H, T \rangle \models \varphi$ to represent that $\langle H, T \rangle$ *satisfies* a formula φ under the recursive conditions:

- $\langle H, T \rangle \not\models \perp$
- $\langle H, T \rangle \models p$ iff $p \in H$
- $\langle H, T \rangle \models \varphi \wedge \psi$ iff $\langle H, T \rangle \models \varphi$ and $\langle H, T \rangle \models \psi$
- $\langle H, T \rangle \models \varphi \vee \psi$ iff $\langle H, T \rangle \models \varphi$ or $\langle H, T \rangle \models \psi$
- $\langle H, T \rangle \models \varphi \rightarrow \psi$ iff both (i) $T \models \varphi \rightarrow \psi$ and (ii) $\langle H, T \rangle \not\models \varphi$ or $\langle H, T \rangle \models \psi$

As usual, we say that $\langle H, T \rangle$ is a *model* of a theory Γ , in symbols $\langle H, T \rangle \models \Gamma$, iff $\langle H, T \rangle \models \varphi$ for all $\varphi \in \Gamma$. It is easy to see that $\langle T, T \rangle \models \Gamma$ iff $T \models \Gamma$ classically. For this reason, we will identify $\langle T, T \rangle$ simply as T and will use ‘ \models ’ indistinctly. By $\text{CL}[\Gamma]$ we denote the set of all classical models of Γ . Interpretation $\langle T, T \rangle = T$ is a *stable (or equilibrium) model* of a theory Γ iff $T \models \Gamma$ and there is no $H \subset T$ such that $\langle H, T \rangle \models \Gamma$. We write $\text{SM}[\Gamma]$ to stand for the set of all stable models of Γ . Note that $\text{SM}[\Gamma] \subseteq \text{CL}[\Gamma]$ by definition.

3 G91 Semantics for Epistemic Theories

In this section we provide a straightforward generalisation of G91 allowing its application to arbitrary modal theories. Formulas are extended with the necessity operator \mathbf{K} according to the following grammar:

$$\varphi ::= \perp \mid a \mid \varphi_1 \wedge \varphi_2 \mid \varphi_1 \vee \varphi_2 \mid \varphi_1 \rightarrow \varphi_2 \mid \mathbf{K} \varphi \quad \text{for any atom } a \in \mathcal{AT}.$$

An (*epistemic*) *theory* is a set of formulas. In our context, the epistemic reading of $\mathbf{K} \psi$ is that “ ψ is one of the agent’s beliefs.” Thus, a formula φ is said to be *subjective* if all its atom occurrences (having at least one) are in the scope of \mathbf{K} . Analogously, φ is said to be *objective* if \mathbf{K} does not occur in φ . For instance, $\neg \mathbf{K} a \vee \mathbf{K} b$ is subjective, $\neg a \vee b$ is objective and $\neg a \vee \mathbf{K} b$ none of the two.

To represent the agent’s beliefs we will use a set \mathbb{W} of propositional interpretations. We call *belief set* to each element $I \in \mathbb{W}$ and *belief view* to the whole set \mathbb{W} . The difference between belief and knowledge is that the former may not hold in the real world. Thus, satisfaction of formulas will be defined with respect to an interpretation $I \subseteq \mathcal{AT}$, possibly $I \notin \mathbb{W}$, that accounts for the real world: the pair (\mathbb{W}, I) is called *belief interpretation* (or interpretation in modal logic KD45). Modal satisfaction is also written $(\mathbb{W}, I) \models \varphi$ (ambiguity is removed by the interpretation on the left) and follows the conditions:

- $(\mathbb{W}, I) \not\models \perp$,
- $(\mathbb{W}, I) \models a$ iff $a \in I$, for any atom $a \in \mathcal{AT}$,
- $(\mathbb{W}, I) \models \psi_1 \wedge \psi_2$ iff $(\mathbb{W}, I) \models \psi_1$ and $(\mathbb{W}, I) \models \psi_2$,
- $(\mathbb{W}, I) \models \psi_1 \vee \psi_2$ iff $(\mathbb{W}, I) \models \psi_1$ or $(\mathbb{W}, I) \models \psi_2$,
- $(\mathbb{W}, I) \models \psi_1 \rightarrow \psi_2$ iff $(\mathbb{W}, I) \not\models \psi_1$ or $(\mathbb{W}, I) \models \psi_2$, and

– $(\mathbb{W}, I) \models \mathbf{K} \psi$ iff $(\mathbb{W}, J) \models \psi$ for all $J \in \mathbb{W}$.

Notice that implication here is classical, that is, $\varphi \rightarrow \psi$ is equivalent to $\neg\varphi \vee \psi$ in this context. A belief interpretation (\mathbb{W}, I) is a *belief model* of Γ iff $(\mathbb{W}, J) \models \varphi$ for all $\varphi \in \Gamma$ and all $J \in \mathbb{W} \cup \{I\}$. We say that \mathbb{W} is an *epistemic model* of Γ , and abbreviate this as $\mathbb{W} \models \Gamma$, iff $(\mathbb{W}, J) \models \varphi$ for all $\varphi \in \Gamma$ and all $J \in \mathbb{W}$. Belief models defined in this way correspond to modal logic KD45 whereas epistemic models correspond to S5.

Example 1. Take the theory $\Gamma_1 = \{\neg\mathbf{K} b \rightarrow a\}$ corresponding to rule (1). An epistemic model $\mathbb{W} \models \Gamma_1$ must satisfy: $\langle \mathbb{W}, J \rangle \models \mathbf{K} b$ or $\langle \mathbb{W}, J \rangle \models a$, for all $J \in \mathbb{W}$. We get three epistemic models from $\mathbf{K} b$, $[\{b\}]$, $[\{a, b\}]$, and $[\{b\}, \{a, b\}]$ and the rest of cases must force a true, so we also get $[\{a\}]$ and $[\{a\}, \{a, b\}]$. In other words, Γ_1 has the same epistemic models as $\mathbf{K} b \vee \mathbf{K} a$. \square

Note that rule (1) alone did not seem to provide any reason for believing b , but we got three epistemic models above satisfying $\mathbf{K} b$. Thus, we will be interested only in some epistemic models (we will call *world views*) that minimize the agent's beliefs in some sense. To define such a minimisation we rely on the following syntactic transformation provided by [21].

Definition 1 (Subjective reduct). *The subjective reduct of a theory Γ with respect to a belief view \mathbb{W} , also written $\Gamma^{\mathbb{W}}$, is obtained by replacing each maximal subformula of the form $\mathbf{K} \varphi$ by: \top , if $\mathbb{W} \models \mathbf{K} \varphi$; by \perp , otherwise. Notice that $\Gamma^{\mathbb{W}}$ is a classical, non-modal theory. \square*

Finally, we impose a fixpoint condition where, depending on whether each belief set $I \in \mathbb{W}$ is required to be a stable model of the reduct or just a classical model, we get G91 or AEL semantics, respectively.

Definition 2 (AEL and G91 world views). *A belief view \mathbb{W} is called an AEL-world view of a theory Γ iff $\mathbb{W} = \text{CL}[\Gamma^{\mathbb{W}}]$, and is called a G91-world view of Γ iff $\mathbb{W} = \text{SM}[\Gamma^{\mathbb{W}}]$. \square*

Example 2 (Example 1 revisited). Take any \mathbb{W} such that $\mathbb{W} \models \mathbf{K} b$. Then, $\Gamma_1^{\mathbb{W}} = \{\perp \rightarrow a\}$ with $\text{CL}[\Gamma_1^{\mathbb{W}}] = [\emptyset, \{a\}, \{b\}, \{a, b\}]$ and $\text{SM}[\Gamma_1^{\mathbb{W}}] = [\emptyset]$. None of the two satisfy $\mathbf{K} b$ so \mathbb{W} cannot be fixpoint for G91 or AEL. If $\mathbb{W} \not\models \mathbf{K} b$ instead, we get $\Gamma_1^{\mathbb{W}} = \{\top \rightarrow a\}$, whose classical models are $\{a\}$ and $\{a, b\}$, but only the former is stable. As a result, $\mathbb{W} = [\{a\}, \{a, b\}]$ is the unique AEL world view and $\mathbb{W} = [\{a\}]$ the unique G91 world view. \square

Example 3. Take now the theory $\Gamma_3 = \{\mathbf{K} a \rightarrow a\}$ corresponding to rule (3). If $\mathbb{W} \models \mathbf{K} a$ we get $\Gamma_3^{\mathbb{W}} = \{\top \rightarrow a\}$ and $\text{CL}[\Gamma_3^{\mathbb{W}}] = \text{SM}[\Gamma_3^{\mathbb{W}}] = \{a\}$ so $\mathbb{W} = [\{a\}]$ is an AEL and G91 world view. If $\mathbb{W} \not\models \mathbf{K} a$, the reduct becomes $\Gamma_3^{\mathbb{W}} = \{\perp \rightarrow a\}$, a classical tautology with unique stable model \emptyset . As a result, $\mathbb{W} = [\emptyset, \{a\}]$ is the other AEL world view, while $\mathbb{W} = [\emptyset]$ is the a second G91 world view. \square

As we can see, the difference between AEL and G91 is that we use classical $\text{CL}[I^{\mathbb{W}}]$ instead of stable $\text{SM}[I^{\mathbb{W}}]$ models, respectively. It is well known that adding the excluded middle axiom $a \vee \neg a$ for all atoms makes equilibrium logic collapse into classical logic. This leads us to the following result.

Theorem 1. \mathbb{W} is an AEL world view of some theory Γ iff \mathbb{W} is a G91-world view of $\Gamma \cup \{a \vee \neg a \mid a \in \mathcal{AT}\}$. \square

4 Founded World Views of Epistemic Specifications

As we explained in the introduction, world view $[\{a\}]$ of $\{\mathbf{K} a \rightarrow a\}$ is considered to be “self-supported” in the literature but, unfortunately, there is no formal definition for such a concept, to the best of our knowledge. To cover this lack, we proceed to extend here the idea of unfounded sets from disjunctive logic programs to the epistemic case. For this purpose, we focus next on the original language of *epistemic specifications* [4] (a fragment of epistemic theories closer to logic programs) on which most approaches have been actually defined.

Let us start by introducing some terminology. An *objective literal* is either an atom $a \in \mathcal{AT}$, its negation $\neg a$ or its double negation $\neg\neg a$. A *subjective literal* is any of the formulas⁴ $\mathbf{K}l$, $\neg\mathbf{K}l$ or $\neg\neg\mathbf{K}l$ where l an objective literal. A *literal* is either an objective or a subjective literal, and is called *negative* if it contains negation and *positive* otherwise. A *rule* is a formula of the form

$$a_1 \vee \dots \vee a_n \leftarrow B_1 \wedge \dots \wedge B_m \quad (4)$$

with $n \geq 0$, $m \geq 0$ and $m + n > 0$, where each a_i is an atom and each B_j is a literal. For any rule r like (4), we define its body as $\text{Body}(r) \stackrel{\text{def}}{=} B_1 \wedge \dots \wedge B_m$ and its head $\text{Head}(r) \stackrel{\text{def}}{=} a_1 \vee \dots \vee a_n$, which we sometimes use as the set of atoms $\{a_1, \dots, a_n\}$. When $n = 0$, $\text{Head}(r) = \perp$ and the rule is a *constraint*, whereas if $m = 0$ then $\text{Body}(r) = \top$ and the rule is a *fact*. The set $\text{Body}_{ob}^+(r)$ collects all atoms occurring in positive objective literals in the body while $\text{Body}_{sub}^+(r)$ collects all atoms occurring in positive subjective literals. An *epistemic specification* or *program* is a set of rules. As with formulas, a program without occurrences of \mathbf{K} is said to be objective (it corresponds to a standard disjunctive logic program with double negation).

Definition 3 (Unfounded set). Let Π be a program and \mathbb{W} a belief view. An unfounded set \mathbb{S} with respect to Π and \mathbb{W} is a non-empty set of pairs where, for each $\langle X, I \rangle \in \mathbb{S}$, we have that X and I are sets of atoms and there is no rule $r \in \Pi$ with $\text{Head}(r) \cap X \neq \emptyset$ satisfying:

1. $(\mathbb{W}, I) \models \text{Body}(r)$
2. $\text{Body}_{ob}^+(r) \cap X = \emptyset$

⁴ We focus here on the study of the operator \mathbf{K} , but epistemic specifications also allow a second operator $\mathbf{M}l$ whose relation to \mathbf{K} is also under debate and, for this reason, we leave it future work.

3. $(\text{Head}(r) \setminus X) \cap I = \emptyset$
4. $\text{Body}_{sub}^+(r) \cap Y = \emptyset$ with $Y = \bigcup \{ X' \mid \langle X', I' \rangle \in \mathbb{S} \}$. □

The definition works in a similar way to standard unfounded sets [12, Definition 3.1]. In fact, the latter corresponds to the first three conditions above, except that we use (\mathbb{W}, I) to check $\text{Body}(r)$, as it may contain now subjective literals. Intuitively, each I represents some potential belief set (or stable model) and X is some set of atoms without a “justifying” rule, that is, there is no $r \in \Pi$ allowing a positive derivation of atoms in X . A rule like that should have a true $\text{Body}(r)$ (condition 1) but not because of positive literals in X (condition 2) and is not used to derive other head atoms outside X (condition 3). The novelty in our definition is the addition of condition 4: to consider r a justifying rule, we additionally require not using any positive literal $\mathbf{K}a$ in the body such that atom a also belongs to any of the unfounded components X' in \mathbb{S} .

Definition 4 (Founded world view). *Let Π be a program and \mathbb{W} be a belief view. We say that \mathbb{W} is unfounded if there is some unfounded-set \mathbb{S} s.t., for every $\langle X, I \rangle \in \mathbb{S}$, we have $I \in \mathbb{W}$ and $X \cap I \neq \emptyset$. \mathbb{W} is called founded otherwise.* □

When Π is an objective program, each pair $\langle X, I \rangle$ corresponds to a standard unfounded set X of some potential stable model I in the traditional sense of [12].

Example 4. Given the single disjunctive rule $a \vee b$ suppose we check the (expected) world view $\mathbb{W} = [\{a\}, \{b\}]$. For $I = \{a\}$ and $X = \{a\}$, rule $a \vee b$ satisfies the four conditions and justifies a . The same happens for $I = \{b\} = X$. So, \mathbb{W} is founded. However, suppose we try with $\mathbb{W}' = [\{a, b\}]$ instead. For $I = \{a, b\}$ we can form $X = \{a\}$ and $X' = \{b\}$ and in both cases, the only rule in the program, $a \vee b$, violates condition 3. As a result, \mathbb{W}' is unfounded due to the set $\mathbb{S}' = \{\langle \{b\}, \{a, b\} \rangle, \langle \{a\}, \{a, b\} \rangle\}$. □

To illustrate how condition 4 works, let us continue with Example 3.

Example 5 (Example 3 continued). Theory $\Gamma_3 = \{\mathbf{K}a \rightarrow a\}$ is also a program. Given belief set $\mathbb{W} = [\{a\}]$ we can observe that $\mathbb{S} = [\langle \{a\}, \{a\} \rangle]$ makes \mathbb{W} unfounded because the unique rule in Γ_3 does not fulfill condition 4: we cannot derive a from a rule that contains $a \in \text{Body}_{sub}^+(r)$. On the other hand, the other G91 world view, $\mathbb{W} = [\emptyset]$, is trivially founded. □

Since Definition 4 only depends on some epistemic program and its selected world views, we can raise it to a general property for any epistemic semantics.

Property 1 (Foundedness). A semantics satisfies *foundedness* when all the world views it assigns to any program Π are founded. □

Approaches proposed after G91 do remove unfounded world views in the examples studied in the literature, but unfortunately, this does not mean that they generally satisfy foundedness. Let us consider a common counterexample.

Example 6. Take the epistemic logic program:

$$a \vee b \qquad a \leftarrow \mathbf{K} b \qquad b \leftarrow \mathbf{K} a \qquad (\Pi_1)$$

whose G91-world views are $\mathbb{W} = [\{a\}, \{b\}]$ and $\mathbb{W}' = [\{a, b\}]$. These are, indeed, the two cases we analysed in Example 4. \mathbb{W} is again founded because $a \vee b$ keeps justifying both possible $\langle X, I \rangle$ pairs, that is, $[\langle \{a\}, \{a\} \rangle]$ and $[\langle \{b\}, \{b\} \rangle]$. However, for \mathbb{W}' we still have the unfounded set $\mathbb{S}' = [\langle \{a\}, \{a, b\} \rangle, \langle \{b\}, \{a, b\} \rangle]$ which violates condition 3 for the first rule as before, but also condition 4 for the other two rules. \square

Note how \mathbb{S}' allows us to spot the root of the derivability problem: to justify a in $\langle \{a\}, \{a, b\} \rangle$ we cannot use $a \leftarrow \mathbf{K} b$ because b is part of the unfounded structure X in the other pair $\langle \{b\}, \{a, b\} \rangle$, and vice versa. Since the variants by Gelfond in [5] (G11) and Kahl et al. [9] (K15) also assign the unfounded world view \mathbb{W}' to Π_1 (in fact, they coincide with G91 for this program), we can conclude that G11 and K15 *do not satisfy foundedness* either.

A more elaborated strategy is adopted by the recent approaches by Fariñas et al. [2] (F15) and Shen and Eiter [20] (S17), that treat the previous world views as candidate solutions⁵, but select the ones with *minimal knowledge* in a second step. This allows removing the unfounded world view $[\{a, b\}]$ in Example 6, because the other solution $[\{a\}, \{b\}]$ provides less knowledge. Unfortunately, this strategy does not suffice to guarantee foundedness, since other formulas (such as constraints) may remove the founded world view, as explained below.

Example 7 (Example 6 continued). Take the program $\Pi_2 = \Pi_1 \cup \{\perp \leftarrow \neg \mathbf{K} a\}$. The constraint rules out $\mathbb{W} = [\{a\}, \{b\}]$ because the latter satisfies $\neg \mathbf{K} a$. In G91, G11, F15 and S17, only world view $\mathbb{W}' = [\{a, b\}]$ is left, so knowledge minimisation has no effect. However, \mathbb{W}' is still unfounded in Π_2 since constraints do not affect that feature (their empty head never justifies any atom). \square

As a conclusion, semantics F15 and S17 *do not satisfy foundedness* either.

5 Founded Autoepistemic Equilibrium Logic

We present now the semantics proposed in this paper, introducing *Founded Autoepistemic Equilibrium Logic* (FAEEL). The basic idea is an elaboration of the belief (or KD45) interpretation (\mathbb{W}, I) already seen but replacing belief sets by HT pairs. Thus, we extend now the idea of *belief view* \mathbb{W} to a non-empty set of HT-interpretations $\mathbb{W} = \{\langle H_1, T_1 \rangle, \dots, \langle H_n, T_n \rangle\}$ and say that \mathbb{W} is *total* when $H_i = T_i$ for all of them, coinciding with the form of belief views $\mathbb{W} = \{T_1, \dots, T_n\}$ we had so far. Similarly, a *belief interpretation* is now redefined as $(\mathbb{W}, \langle H, T \rangle)$, or simply (\mathbb{W}, H, T) , where \mathbb{W} is a belief view and $\langle H, T \rangle$ stands for the real world, possibly not in \mathbb{W} . Next, we redefine the satisfaction relation from a combination of modal logic KD45 and HT. A belief interpretation $\mathcal{I} = (\mathbb{W}, H, T)$ satisfies a formula φ , written $\mathcal{I} \models \varphi$, iff:

⁵ In [2], these candidate world views are called *epistemic equilibrium models* while selected world views receive the name of *autoepistemic equilibrium models*.

- $\mathcal{I} \not\models \perp$,
- $\mathcal{I} \models a$ iff $a \in H$, for any atom $a \in \mathcal{AT}$,
- $\mathcal{I} \models \psi_1 \wedge \psi_2$ iff $\mathcal{I} \models \psi_1$ and $\mathcal{I} \models \psi_2$,
- $\mathcal{I} \models \psi_1 \vee \psi_2$ iff $\mathcal{I} \models \psi_1$ or $\mathcal{I} \models \psi_2$,
- $\mathcal{I} \models \psi_1 \rightarrow \psi_2$ iff both: (i) $\mathcal{I} \not\models \psi_1$ or $\mathcal{I} \models \psi_2$; and (ii) $(\mathbb{W}^t, T) \not\models \psi_1$ or $(\mathbb{W}^t, T) \models \psi_2$, where $\mathbb{W}^t = \{T_i \mid \langle H_i, T_i \rangle \in \mathbb{W}\}$.
- $\mathcal{I} \models \mathbf{K}\psi$ iff $(\mathbb{W}, H_i, T_i) \models \psi$ for all $\langle H_i, T_i \rangle \in \mathbb{W}$.

For total belief interpretations, this new satisfaction relation collapses to the one in Sect. 3 (that is, KD45). Interpretation (\mathbb{W}, H, T) is a *belief model* of Γ iff $(\mathbb{W}, H_i, T_i) \models \varphi$ for all $\langle H_i, T_i \rangle \in \mathbb{W} \cup \{\langle H, T \rangle\}$ and all $\varphi \in \Gamma$ – additionally, when $\langle H, T \rangle \in \mathbb{W}$, we further say that \mathbb{W} is an *epistemic model* of Γ , abbreviated as $\mathbb{W} \models \Gamma$.

Proposition 1 (Persistence). $(\mathbb{W}, H, T) \models \varphi$ implies $(\mathbb{W}^t, T) \models \varphi$. □

A belief model just captures collections of HT models which need not be in equilibrium. To make the agent’s beliefs correspond to stable models we impose a particular minimisation criterion on belief models.

Definition 5. We define the partial order $\mathcal{I}' \preceq \mathcal{I}$ for belief interpretations $\mathcal{I}' = (\mathbb{W}', H', T')$ and $\mathcal{I} = (\mathbb{W}, H, T)$ when the following three conditions hold:

1. $T' = T$ and $H' \subseteq H$, and
2. for every $\langle H_i, T_i \rangle \in \mathbb{W}$, there is some $\langle H'_i, T_i \rangle \in \mathbb{W}'$, with $H'_i \subseteq H_i$.
3. for every $\langle H'_i, T_i \rangle \in \mathbb{W}'$, there is some $\langle H_i, T_i \rangle \in \mathbb{W}$, with $H'_i \subseteq H_i$. □

As usual, $\mathcal{I}' \prec \mathcal{I}$ means $\mathcal{I}' \preceq \mathcal{I}$ and $\mathcal{I}' \neq \mathcal{I}$. The intuition for $\mathcal{I}' \preceq \mathcal{I}$ is that \mathcal{I}' contains less information than \mathcal{I} for each fixed T_i component. As a result, $\mathcal{I}' \models \varphi$ implies $\mathcal{I} \models \varphi$ for any formula φ without implications other than $\neg\psi = \psi \rightarrow \perp$.

Definition 6. A total belief interpretation $\mathcal{I} = (\mathbb{W}, T)$ is said to be an equilibrium belief model of some theory Γ iff \mathcal{I} is a belief model of Γ and there is no other belief model \mathcal{I}' of Γ such that $\mathcal{I}' \prec \mathcal{I}$. □

By $\text{EQB}[\Gamma]$ we denote the set of equilibrium belief models of Γ . As a final step, we impose a fixpoint condition to minimise the agent’s knowledge as follows.

Definition 7. A belief view \mathbb{W} is called an equilibrium world view of Γ iff:

$$\mathbb{W} = \{ T \mid (\mathbb{W}, T) \in \text{EQB}[\Gamma] \} \quad \square$$

Example 8 (Example 5 continued). Back to $\Gamma_3 = \{\mathbf{K}a \rightarrow a\}$, remember its unique founded G91-world view was $[\emptyset]$. It is easy to see that $\mathcal{I} = ([\emptyset], \emptyset) \in \text{EQB}[\Gamma_3]$ because $([\emptyset], \emptyset) \models \Gamma_3$ and no smaller belief model can be obtained. Moreover, $[\emptyset]$ is an equilibrium world view of Γ_3 since no other $T \notin [\emptyset]$ satisfies $([\emptyset], T) \in \text{EQB}[\Gamma_3]$. The only possibility is $([\emptyset], \{a\})$ but it fails because there is a smaller belief model $([\emptyset], \emptyset, \{a\})$ satisfying $\mathbf{K}a \rightarrow a$. As for the other potential world view $[\{a\}]$, it is not in equilibrium: we already have $\mathcal{I}' = ([\{a\}], \{a\}) \notin \text{EQB}[\Gamma_3]$ because the smaller interpretation $\mathcal{I}'' = ([\emptyset, \{a\}], \{a\}, \{a\})$ also satisfies Γ_3 . In particular, note that $\mathcal{I}'' \not\models \mathbf{K}a$ and, thus, clearly satisfies $\mathbf{K}a \rightarrow a$. □

The logic induced by equilibrium world views is called *Founded Autoepistemic Equilibrium Logic* (FAEEL). A first important property is:

Theorem 2. *FAEEL satisfies foundedness.* □

A second interesting feature is that equilibrium world views are also G91-world views though the converse may not be the case (as we just saw in Example 8). This holds, not only for programs, but in general for any theory:

Theorem 3. *For any theory Γ , its equilibrium world views are also G91-world views of Γ .* □

In other words, FAEEL is strictly stronger than G91, something that, as we see next, is not the case in other approaches in the literature.

Example 9. The following program:

$$a \vee b \qquad c \leftarrow \mathbf{K} a \qquad \perp \leftarrow \neg c \qquad (\Pi_3)$$

has no G91-world views, but according to G11, K15, F15 and S17 has world view $[\{a, c\}]$. This example was also used in [1] to show that these semantics do not satisfy another property, called there *epistemic splitting*. □

Example 10 (Example 6 continued). Take again program Π_1 whose G91-world views were $\mathbb{W} = [\{a\}, \{b\}]$ and $\mathbb{W}' = [\{a, b\}]$. Since \mathbb{W}' is unfounded, it cannot be an equilibrium world view (Theorem 2), leaving \mathbb{W} as the only candidate (Theorem 3). Let us check that this is in fact an equilibrium world view. First, note that $\mathcal{I} = ([\{a\}, \{b\}], \{a\}) \in \text{EQB}[\Pi_1]$ because there is no model \mathcal{I}' of Π_1 such that $\mathcal{I}' \preceq \mathcal{I}$. In fact, it is easy to see that $([\langle H_1, \{a\} \rangle, \langle H_2, \{b\} \rangle], H_3, \{a\})$ is not a model of the rule $a \vee b$ if $H_i = \emptyset$ for any $i \in \{1, 2, 3\}$. Symmetrically, we have that $\mathcal{I}' = ([\{a\}, \{b\}], \{b\}) \in \text{EQB}[\Pi_1]$ too. Finally, we have to check that no other $T \notin [\{a\}, \{b\}]$ can form an equilibrium belief model. For the case $T = \emptyset$, it is easy to check that $([\{a\}, \{b\}], \emptyset)$ does not satisfy $a \vee b$. For $T = \{a, b\}$, we have that $\mathcal{I}'' = ([\{a\}, \{b\}], \{a, b\}) \notin \text{EQB}[\Pi_1]$ because, for instance, the smaller $\mathcal{I}''' = ([\{a\}, \{b\}], \{a\}, \{a, b\})$ is a model of Π_1 . □

Theorems 2 and 3 assert that any equilibrium world view is a founded G91-world view. The natural question is whether the opposite also holds. In Examples 8, 9 and 10 we did not find any counterexample, and this is in fact a general property, as stated below.

Main Theorem. *Given any program Π , its equilibrium world views coincide with its founded G91-world views.* □

An interesting observation is that in all the original examples of epistemic specifications [4, 7] used by Gelfond to introduce G91, modal operators occurred in the scope of negation. Negated beliefs never incur unfoundedness, so this feature could not be spotted using this family of examples. In fact, under this syntactic restriction, FAEEL and G91 coincide.

program	world views	program	G91/G11/FAEEL	K15/F15/S17
$a \vee b$	$[\{a\}, \{b\}]$	$a \leftarrow \neg \mathbf{K} \neg a$	$[\emptyset], [\{a\}]$	$[\{a\}]$
$a \vee b$ $a \leftarrow \mathbf{K} b$	$[\{a\}, \{b\}]$	$a \vee b$ $a \leftarrow \neg \mathbf{K} \neg b$	none	$[\{a\}]$
$a \vee b$ $a \leftarrow \neg \mathbf{K} b$	$[\{a\}]$	$a \vee b$ $a \leftarrow \mathbf{K} \neg b$	$[\{a\}], [\{a\}, \{b\}]$	$[\{a\}, \{b\}]$
$a \vee b$ $c \leftarrow \neg \mathbf{K} b$	$[\{a, c\}, \{b, c\}]$	$a \leftarrow b$ $b \leftarrow \neg \mathbf{K} \neg a$	$[\emptyset], [\{a, b\}]$	$[\{a, b\}]$
$a \leftarrow \neg \mathbf{K} b$ $b \leftarrow \neg \mathbf{K} a$	$[\{a\}], [\{b\}]$	$a \leftarrow \neg \mathbf{K} \neg b$ $b \leftarrow \neg \mathbf{K} \neg a$	$[\emptyset], [\{a\}, \{b\}]$	$[\{a\}, \{b\}]$
$a \leftarrow \neg \mathbf{K} \neg a$ $a \leftarrow \neg \mathbf{K} a$	$[\{a\}]$			

Fig. 1. On the left, examples where G91, G11, K15, F15, S17 and FAEEL agree. On the right, examples where FAEEL/G91/G11 differ from K15/F15/S17.

Proposition 2. *For any theory where all occurrences of \mathbf{K} are in the scope of negation, we have that the equilibrium world views and the G91-world views coincide.* \square

Proposition 2 also holds for semantics [21, 22] that are conservative extensions of G91, as well as for G11. Apart from foundedness, [1] recently proposed other four properties for semantics of epistemic specifications. We analyse here three of them, omitting the so-called *epistemic splitting* due to lack of space.

1. *supra-ASP* holds when, for any objective theory Γ , either Γ has a unique world view $\mathbb{W} = \text{SM}[\Gamma] \neq \emptyset$ or $\text{SM}[\Gamma] = \emptyset$ and Γ has no world view.
2. *supra-S5* holds when every world view \mathbb{W} of a theory Γ is also an S5-model of Γ (that is, $\mathbb{W} \models \Gamma$).
3. *subjective constraint monotonicity* holds when, for any theory Γ and any subjective constraint $\perp \leftarrow \varphi$, we have that \mathbb{W} is a world view of $\Gamma \cup \{\perp \leftarrow \varphi\}$ iff both \mathbb{W} is a world view of Γ and \mathbb{W} is not an S5-model of φ .

Proposition 3. *FAEEL satisfies supra-ASP, supra-S5 and subjective constraint monotonicity.* \square

All semantics discussed in this paper satisfy the above first two properties but most of them fail for subjective constraint monotonicity, as first discussed in [10]. In fact, a variation of Example 9 can be used to show that K15, F15 and S17 do not satisfy this property.

Example 11 (Example 9 continued). Suppose we remove the constraint (last rule) from Π_3 getting the program $\Pi_4 = \{a \vee b, c \leftarrow \mathbf{K} a\}$. All semantics, including G91 and FAEEL, agree that Π_4 has a unique world view $[\{a\}, \{b\}]$. Suppose we add now a subjective constraint $\Pi_5 = \Pi_4 \cup \{\perp \leftarrow \neg \mathbf{K} c\}$. This addition leaves G91 and FAEEL without world views (due to subjective constraint monotonicity) the same happens for G11, but not for K15, F15 and S17, which provide a *new* world view $[\{a, c\}]$ not obtained before adding the subjective constraint. \square

Tables 1 and 2 show a list of examples taken from Table 4 in [2] and their world views according to different semantics.

program	G91	G11/FAEEL	K15	F15/S17
$a \leftarrow \neg \mathbf{K} \neg b \wedge \neg b$ $b \leftarrow \neg \mathbf{K} \neg a \wedge \neg a$		$[\emptyset], [\{a\}, \{b\}]$		$[\{a\}, \{b\}]$
$a \leftarrow \mathbf{K} a$	$[\emptyset], [\{a\}]$		$[\emptyset]$	
$a \leftarrow \mathbf{K} a$ $a \leftarrow \neg \mathbf{K} a$	$[\{a\}]$		none	

Fig. 2. Examples splitting different semantics. Examples 6 and 9 in the paper can be used to further split FAEEL and G11.

6 Conclusions

In order to characterise self-supported world-views, already present in Gelfond’s 1991 semantics [4] (G91), we have extended the definition of unfounded sets from standard logic programs to epistemic specifications. As a result, we proposed the *foundedness* property for epistemic semantics, which is not satisfied by other approaches in the literature. Our main contribution has been the definition of a new semantics, based on the so-called *Founded Autoepistemic Equilibrium Logic* (FAEEL), that satisfies foundedness. This semantics actually covers the syntax of any arbitrary modal theory and is a combination of Equilibrium Logic and Autoepistemic Logic. As a main result, we were able to prove that, for the syntax of epistemic specifications, FAEEL world views coincide with the set of G91 world views that are founded. We showed how this semantics behaves on a set of common examples in the literature and proved that it satisfies other three basic properties: all world views are S5 models (*supra-S5*); standard programs have (at most) a unique world view containing all the stable models (*supra-ASP*); and subjective constraints just remove world views (monotonicity). FAEEL also satisfies the property of *epistemic splitting* as proposed in [1], but we leave the proof and discussion for future work, together with a formal comparison with other approaches.

Acknowledgements. We are thankful to Michael Gelfond and David Pearce for their helpful comments on early versions of this work and to the anonymous reviewers for their valuable feedback in improving the paper.

References

1. Cabalar, P., Fandinno, J., Fariñas del Cerro, L.: Splitting Epistemic Logic Programs. arXiv e-prints [arXiv:1812.08763](https://arxiv.org/abs/1812.08763). December 2018
2. Fariñas del Cerro, L., Herzig, A., Su, E.I.: Epistemic equilibrium logic. In: Proceedings of the International Joint Conference on Artificial Intelligence, IJCAI 2015, pp. 2964–2970. AAAI Press (2015)
3. Gelder, A.V., Ross, K.A., Schlipf, J.S.: The well-founded semantics for general logic programs. *J. ACM* **38**(3), 620–650 (1991)
4. Gelfond, M.: Strong introspection. In: Dean, T.L., McKeown, K. (eds.) Proceedings of the AAAI Conference, vol. 1, pp. 386–391. AAAI Press/The MIT Press (1991)
5. Gelfond, M.: New semantics for epistemic specifications. In: Delgrande, J.P., Faber, W. (eds.) LPNMR 2011. LNCS (LNAI), vol. 6645, pp. 260–265. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-20895-9_29
6. Gelfond, M., Lifschitz, V.: The stable model semantics for logic programming. In: Proceedings of the 5th International Conference on Logic Programming, ICLP 1988, pp. 1070–1080 (1988)
7. Gelfond, M., Przymusinska, H.: Reasoning on open domains. In: Proceedings of the International Conference on Logic Programming and Non-Monotonic Reasoning, LPNMR 1993, pp. 397–413. MIT Press (1993)
8. Heyting, A.: Die formalen Regeln der intuitionistischen Logik. *Sitzungsberichte der Preussischen Akademie der Wissenschaften, Physikalisch-mathematische Klasse*, pp. 42–56 (1930)
9. Kahl, P., Watson, R., Balai, E., Gelfond, M., Zhang, Y.: The language of epistemic specifications (refined) including a prototype solver. *J. Logic Comput.* (2015). (Special issue article)
10. Kahl, P.T., Leclerc, A.P.: Epistemic logic programs with world view constraints. In: ICLP (Technical Communications), vol. 64, pp. 1:1–1:17. OASICS, Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik (2018)
11. Konolige, K.: On the relation between default and autoepistemic logic. *Artif. Intell.* **35**(3), 343–382 (1988)
12. Leone, N., Rullo, P., Scarcello, F.: Disjunctive stable models: unfounded sets, fix-point semantics, and computation. *Inf. Comput.* **135**(2), 69–112 (1997)
13. Marek, V.W., Truszczynski, M.: Relating autoepistemic and default logics. In: KR, pp. 276–288. Morgan Kaufmann (1989)
14. Moore, R.C.: Possible-world semantics for autoepistemic logic. In: Proceedings of the Non-Monotonic Reasoning Workshop, Mohonk Mountain House, New Paltz, NY 12561, USA, 17–19 October 1984. pp. 344–354. American Association for Artificial Intelligence (AAAI) (1984)
15. Moore, R.C.: Semantical considerations on nonmonotonic logic. *Artif. Intell.* **25**(1), 75–94 (1985). [https://doi.org/10.1016/0004-3702\(85\)90042-6](https://doi.org/10.1016/0004-3702(85)90042-6)
16. Niemelä, I.: Constructive tightly grounded autoepistemic reasoning. In: IJCAI, pp. 399–405. Morgan Kaufmann (1991)
17. Pearce, D.: A new logical characterisation of stable models and answer sets. In: Dix, J., Pereira, L.M., Przymusinski, T.C. (eds.) NMELP 1996. LNCS, vol. 1216, pp. 57–70. Springer, Heidelberg (1997). <https://doi.org/10.1007/BFb0023801>
18. Schwarz, G.: Autoepistemic logic of knowledge. In: Nerode, A., Marek, V.W., Subrahmanian, V.S. (eds.) Logic Programming and Non-monotonic Reasoning, Proceedings of the First International Workshop, Washington, D.C., USA, July 1991, pp. 260–274. The MIT Press (1991)

19. Schwarz, G.: Minimal model semantics for nonmonotonic modal logics. In: Proceedings of the Seventh Annual IEEE Symposium on Logic in Computer Science 1992, LICS 1992, pp. 34–43. IEEE (1992)
20. Shen, Y., Eiter, T.: Evaluating epistemic negation in answer set programming (extended abstract). In: Proceedings of the International Joint Conference on Artificial Intelligence, IJCAI 2017, pp. 5060–5064 (2017)
21. Truszczyński, M.: Revisiting epistemic specifications. In: Balduccini, M., Son, T.C. (eds.) Logic Programming, Knowledge Representation, and Nonmonotonic Reasoning. LNCS (LNAI), vol. 6565, pp. 315–333. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-20832-4_20
22. Wang, K., Zhang, Y.: Nested epistemic logic programs. In: Baral, C., Greco, G., Leone, N., Terracina, G. (eds.) LPNMR 2005. LNCS (LNAI), vol. 3662, pp. 279–290. Springer, Heidelberg (2005). https://doi.org/10.1007/11546207_22