Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI 2015)

Epistemic Equilibrium Logic*

Luis Fariñas del Cerro, Andreas Herzig and Ezgi Iraz Su
University of Toulouse, CNRS, IRIT
Toulouse, France
http://www.irit.fr

Abstract

We add epistemic modal operators to the language of here-and-there logic and define epistemic here-and-there models. We then successively define epistemic equilibrium models and autoepistemic equilibrium models. The former are obtained from here-and-there models by the standard minimisation of truth of Pearce's equilibrium logic; they provide an epistemic extension of that logic. The latter are obtained from the former by maximising the set of epistemic possibilities; they provide a new semantics for Gelfond's epistemic specifications. For both definitions we characterise strong equivalence by means of logical equivalence in epistemic here-and-there logic.

1 Introduction

Gelfond extended disjunctive answer set programs by epistemic operators in order to quantify over answer sets [Gelfond, 1991; 1994; Baral and Gelfond, 1994]. He called programs with epistemic operators in rule bodies *epistemic specifications*. They allow to reason about incomplete information, understood as situations where there are multiple answer sets. The semantics of epistemic specifications is in terms of *world views*, which are maximal collections of answer sets. Similar to answer set semantics, a world view $\mathcal{S}$ of an epistemic specification Π is defined by means of the reduct $\Pi^{\mathcal{S}}$ of Π by $\mathcal{S}$. The construction is in two steps: first, compute the reduct $\Pi^{\mathcal{S}}$, eliminating the modal operators; second, compute the maximal collection of answer sets of $\Pi^{\mathcal{S}}$ and check whether it equals $\mathcal{S}$.

Subsequently various different semantics were proposed for epistemic specifications [Chen, 1997; Wang and Zhang, 2005; Truszczyński, 2011; Gelfond, 2011; Kahl, 2014]. Some propose other definitions of reducts, while others propose semantics inspired by the Kripke semantics of modal logics. We here introduce a new semantics that is based on an epistemic extension of Pearce's equilibrium logic [1996; 2006]. This had been undertaken by Wang and Zhang [2005], however for the somewhat outdated version of epistemic specifications of [Gelfond, 1991; 1994]. Our version is closer to Gelfond's [2011] and Kahl's [2014] more recent versions.

Pearce's equilibrium logic allows to characterise answer set semantics of non-epistemic logic programs. The underlying here-and-there logic (HT) characterises their *strong equivalence* [Lifschitz *et al.*, 2001]. We here add two epistemic operators K and $\hat{\mathsf{K}}$ to the language of HT and define epistemic HT models (EHT models). We then define epistemic equilibrium models by generalising the usual minimality criterion over HT models to EHT models. We finally define autoepistemic equilibrium models (AEEMs) as epistemic equilibrium models that are maximal both under set inclusion and under a preference ordering on epistemic equilibrium models. The latter provide a new logical semantics for epistemic specifications and more generally for nested epistemic logic programs. We establish a strong equivalence result and compare our semantics with existing approaches. We demonstrate by means of examples that all other semantics differ from our approach and that Kahl's comes closest.

The rest of the paper is organised as follows. Section 2 recalls epistemic specifications. Section 3 introduces the logic EHT. Section 4 defines epistemic equilibrium models (EEM) and autoepistemic equilibrium models (AEEM). Section 5 compares the latter with existing semantics of epistemic specifications, in particular with Kahl's. Section 6 provides strong equivalence characterisations in terms of EHT equivalence for both kinds of models. Section 7 concludes.

2 Epistemic specifications

Several versions of epistemic specifications were introduced by Gelfond [1991; 2011]. Kahl [2014] recently proposed a further improvement that we recall here.

2.1 The language of epistemic specifications

The language of epistemic specifications extends that of disjunctive logic programming by the modal operator K. The formula $\mathsf{K}\varphi$ is read "φ is known to be true".[1] Literals of this language are of three different kinds: *objective literals (l)*,

*We thank Pedro Cabalar, Michael Gelfond, Patrick Thor Kahl and David Pearce for useful discussions and feedback. Our work was supported by the *French-Spanish Lab of Advanced Studies in Information Representation and Processing* (LEA CNRS).

[1]The original presentation also has a modal operator M that is dual to K: instead of $\mathsf{K}\,\mathtt{not}\,l$ it has $\mathtt{not}\,\mathsf{M}\,l$ and instead of $\mathtt{not}\,\mathsf{K}\,\mathtt{not}\,l$ it has $\mathsf{M}\,l$. Due to this definability via K we do without M here.

literal L	if $\mathcal{S} \models L$	if $\mathcal{S} \not\models L$
$\mathsf{K}\,\lambda$	replace by λ	replace by $\bot$
$\mathtt{not}\,\mathsf{K}\,\lambda$	replace by $\top$	replace by $\mathtt{not}\,\lambda$

Table 1: Kahl's definition of reduct.

extended objective literals (λ) and *extended subjective literals (L)*. They are defined by the following grammar:

$$l ::= p \mid {\sim}p$$
$$\lambda ::= l \mid \mathtt{not}\,l$$
$$L ::= \mathsf{K}\,\lambda \mid \mathtt{not}\,\mathsf{K}\,\lambda$$

where p ranges over the set of propositional variables $\mathbb{P}$. So the language has two negations: strong negation $\sim$ and default negation $\mathtt{not}$, where $\mathtt{not}\,\varphi$ is read "φ is false by default".[2]

A *rule* ρ is of the form

$$l_1\ \mathtt{or}\ \ldots\ \mathtt{or}\ l_m \ \leftarrow\ G_1\ ,\ \ldots\ ,\ G_n$$

where the literals of the head $l_1\ \mathtt{or}\ \ldots\ \mathtt{or}\ l_m$ are objective literals and the literals of the body $G_1, \ldots, G_n$ are extended literals (either extended objective literals or extended subjective literals). We suppose that the head of ρ is $\bot$ if $m = 0$ and that the body of ρ is $\top$ if $n = 0$.

An *epistemic specification* is a finite collection of rules. Here is Gelfond's [1991] 'eligibility' example that was taken up in almost all papers on epistemic specifications:

$$\begin{aligned} h\ \mathtt{or}\ f &\leftarrow \\ e &\leftarrow h \\ e &\leftarrow f, m \\ {\sim}e &\leftarrow {\sim}h, {\sim}f \\ i &\leftarrow \mathtt{not}\,\mathsf{K}\,e, \mathtt{not}\,\mathsf{K}\,{\sim}e \end{aligned}$$

where h stands for 'high GPA', f stands for 'fair GPA', e stands for 'eligible for scholarship' and i stands for 'to be interviewed'. Let us call that epistemic specification Π_G.

2.2 The semantics of epistemic specifications

Let S be a consistent set of objective literals, i.e., a set of objective literals such that there is no $p \in \mathbb{P}$ with both $p \in S$ and ${\sim}p \in S$. Satisfaction of extended objective literals in S is defined by:

$$\begin{array}{lll} S \models_{\mathsf{ES}} l & \text{if} & l \in S; \\ S \models_{\mathsf{ES}} \mathtt{not}\,l & \text{if} & l \notin S. \end{array}$$

Let $\mathcal{S}$ be a non-empty collection of such consistent S's. Satisfaction of extended subjective literals in $\mathcal{S}$ is defined by:

$$\begin{array}{lll} \mathcal{S} \models_{\mathsf{ES}} \mathsf{K}\,\lambda & \text{if} & S \models_{\mathsf{ES}} \lambda \text{ for every } S \in \mathcal{S}; \\ \mathcal{S} \models_{\mathsf{ES}} \mathtt{not}\,\mathsf{K}\,\lambda & \text{if} & \mathcal{S} \not\models_{\mathsf{ES}} \mathsf{K}\,\lambda. \end{array}$$

For example, $\{\{p\}\} \models_{\mathsf{ES}} \mathtt{not}\,\mathsf{K}\,\mathtt{not}\,p$ because $\{\{p\}\} \not\models_{\mathsf{ES}} \mathsf{K}\,\mathtt{not}\,p$.

Let Π be an epistemic specification and let $\mathcal{S}$ be a non-empty collection of consistent sets of objective literals. Whether $\mathcal{S}$ is a world view of Π is decided by computing the reduct of Π and then checking a fixed-point equation. The reduct $\rho^{\mathcal{S}}$ of a rule $\rho \in \Pi$ with respect to $\mathcal{S}$ is obtained from ρ by eliminating the K operator according to Table 1. The *reduct* of Π with respect to $\mathcal{S}$, noted $\Pi^{\mathcal{S}}$, is then defined as

$$\Pi^{\mathcal{S}} = \{\rho^{\mathcal{S}} \ :\ \rho \in \Pi\}.$$

Finally, $\mathcal{S}$ is a *world view* of Π if it equals the set of all answer sets of $\Pi^{\mathcal{S}}$. As the case $\mathtt{not}\,\mathsf{K}\,\mathtt{not}\,l$ may introduce double negation $\mathtt{not}\,\mathtt{not}\,l$, the computation of such answer sets has to resort to answer-set programming with nested expressions [Lifschitz *et al.*, 1999].

For example, the only world view of Gelfond's 'eligibility' specification Π_G of Section 2.1 is $\{\{h, e, i\}, \{f, i\}\}$. Table 2 contains more examples. The reduct of $\{p \leftarrow \mathsf{K}\,p\}$ by $\{\emptyset\}$ is $p \leftarrow \bot$ and its reduct by $\{\{p\}\}$ is $p \leftarrow p$. Both have exactly one answer set: $\{\emptyset\}$. Only the former matches its reduct, so the unique world view is $\{\emptyset\}$. The reduct of $\{p\ \mathtt{or}\ q \leftarrow, p \leftarrow \mathtt{not}\,\mathsf{K}\,\mathtt{not}\,q\}$ by $\{\{p\}\}$ is $\{p\ \mathtt{or}\ q \leftarrow, p \leftarrow \mathtt{not}\,\mathtt{not}\,q\}$ and its reduct by $\{\{q\}\}$ is $\{p\ \mathtt{or}\ q \leftarrow, p \leftarrow \top\}$. Both have exactly one answer set: $\{p\}$; therefore the former $\{\{p\}\}$ is the unique world view.[3]

Observe that when Π does not contain K then it has at most one world view: the collection of all answer sets of Π.

formula	world views
$\{p \leftarrow \mathsf{K}\,p\}$	$\{\emptyset\}$
$\{p \leftarrow \mathtt{not}\,\mathsf{K}\,\mathtt{not}\,p\}$	$\{\{p\}\}$
$\{p\ \mathtt{or}\ q \leftarrow\ ,\ \ p \leftarrow \mathtt{not}\,\mathsf{K}\,\mathtt{not}\,q\}$	$\{\{p\}\}$
$\{p\ \mathtt{or}\ q \leftarrow\ ,\ \ r \leftarrow \mathsf{K}\,p\}$	$\{\{p\}, \{q\}\}$

Table 2: Examples of epistemic specifications and their world views.

3 Epistemic here-and-there logic EHT

The logic of here-and-there HT is a monotonic three-valued logic that is intermediate between classical logic and intuitionistic logic. An HT model is an ordered pair (H, T) of valuations (sets of propositional variables) satisfying the *heredity constraint* $H \subseteq T$. In this section we introduce epistemic HT logic (EHT), which extends HT by two epistemic modal operators K and $\hat{\mathsf{K}}$ in the spirit of intuitionistic modal logics [Fischer-Servi, 1976; Fariñas del Cerro and Raggio, 1983; Simpson, 1994; Bierman and de Paiva, 2000]. Its epistemic HT models (EHT models) are collections of HT models. From the perspective of modal logic, an EHT model can be viewed as a refinement of S5 models (which are sets of valuations) where valuations are replaced by HT models.

3.1 The language of EHT

The language of EHT is given by the following grammar:

$$\varphi ::= p \mid \bot \mid \varphi \wedge \varphi \mid \varphi \vee \varphi \mid \varphi \rightarrow \varphi \mid \mathsf{K}\varphi \mid \hat{\mathsf{K}}\varphi$$

where p ranges over the set of propositional variables $\mathbb{P}$. An EHT *theory* is a finite set of EHT formulas, noted $\Phi, \Psi, \ldots$

[2] Gelfond's original language is slightly different, with ${\sim}\mathsf{K}\,l$ instead of $\mathtt{not}\,\mathsf{K}\,l$, etc. They however have the same semantics, *mutatis mutandis*.

[3] Gelfond's [2011] approach differs for the case $\mathtt{not}\,\mathsf{K}\,\lambda$: when $\mathcal{S} \not\models \mathtt{not}\,\mathsf{K}\,\lambda$ then it is replaced by $\bot$. As a result of this subtle difference, Gelfond gets the world views $\{\emptyset\}$ and $\{\{p\}\}$ for the 2nd example, and no world view at all for the 3rd. As Kahl argues, both are not as intuitive as his results.

The set of propositional variables occurring in a formula φ is noted $\mathbb{P}_\varphi$. For example, $\mathbb{P}_{\mathsf{K}(p\to q)} = \{p, q\}$. This generalises to EHT theories: $\mathbb{P}_\Phi = \bigcup_{\varphi\in\Phi} \mathbb{P}_\varphi$.

As usual, $\top$, $\neg\varphi$ and $\varphi \leftrightarrow \psi$ respectively abbreviate $\bot\to\bot$, $\varphi\to\bot$ and $(\varphi\to\psi) \wedge (\psi \to \varphi)$. A formula is said to be *non-modal* if it does not contain the modal operators K and $\hat{\mathsf{K}}$.

3.2 EHT models

An EHT model is an ordered pair $(\mathcal{T}, \hbar)$ in which

- $\mathcal{T} \subseteq 2^{\mathbb{P}}$ is a nonempty set of valuations;
- $\hbar : \mathcal{T} \to 2^{\mathbb{P}}$ is a map such that $\hbar(T) \subseteq T$ for every $T \in \mathcal{T}$.

When $\mathcal{T} = \{T\}$ is a singleton then the EHT model $(\mathcal{T}, \hbar)$ can be identified with the HT model $(\hbar(T), T)$.

The inclusion constraint on $\hbar$ generalises the *heredity* constraint of HT logic to EHT. We say that $(\mathcal{T}, \hbar)$ is *total on* $\mathcal{S} \subseteq \mathcal{T}$ if $\hbar(T) = T$ for every $T \in \mathcal{S}$. When $(\mathcal{T}, \hbar)$ is total on $\mathcal{T}$ then $\hbar$ is the identity function *id*. We identify $(\mathcal{T}, id)$ with the classical $\mathsf{S5}$ model $\mathcal{T}$.

A *pointed* EHT *model* is a pair $((\mathcal{T}, \hbar), T_0)$ where $(\mathcal{T}, \hbar)$ is an EHT model and $T_0 \in \mathcal{T}$, the latter being the actual world. A *multipointed* EHT *model* is a pair $((\mathcal{T}, \hbar), \mathcal{T}_0)$ where $(\mathcal{T}, \hbar)$ is an EHT model and $\mathcal{T}_0 \subseteq \mathcal{T}$.

3.3 EHT truth conditions

We now define the satisfaction relation for EHT formulas. Those for $\bot$, $\wedge$ and $\vee$ are standard.

$$\begin{array}{lll}
(\mathcal{T},\hbar), T \models_{\mathsf{EHT}} p & \text{if } p \in \hbar(T) \\
(\mathcal{T},\hbar), T \models_{\mathsf{EHT}} \varphi\to\psi & \text{if } (\mathcal{T},\hbar), T \models_{\mathsf{EHT}} \varphi \Rightarrow (\mathcal{T},\hbar), T \models_{\mathsf{EHT}} \psi \text{ and} \\
 & (\mathcal{T},id), T \models_{\mathsf{EHT}} \varphi \Rightarrow (\mathcal{T},id), T \models_{\mathsf{EHT}} \psi \\
(\mathcal{T},\hbar), T \models_{\mathsf{EHT}} \mathsf{K}\varphi & \text{if } (\mathcal{T},\hbar), T' \models_{\mathsf{EHT}} \varphi \text{ for every } T' \in \mathcal{T} \\
(\mathcal{T},\hbar), T \models_{\mathsf{EHT}} \hat{\mathsf{K}}\varphi & \text{if } (\mathcal{T},\hbar), T' \models_{\mathsf{EHT}} \varphi \text{ for some } T' \in \mathcal{T}
\end{array}$$

It follows that for negation we have
$(\mathcal{T},\hbar), T \models_{\mathsf{EHT}} \neg\varphi$ iff $(\mathcal{T},\hbar), T \not\models_{\mathsf{EHT}} \varphi$ and $(\mathcal{T},id), T \not\models_{\mathsf{EHT}} \varphi$.
This can be simplified further, see Proposition 3.5.1 below. For example, $(\mathcal{T},\hbar), T_0 \models_{\mathsf{EHT}} p\vee\neg p$ iff $p \in \hbar(T_0)$ or $p \notin T_0$. Raising the bar of our examples,

$$(\mathcal{T},\hbar), T_0 \models_{\mathsf{EHT}} \mathsf{K}(p\vee\neg p) \to \hat{\mathsf{K}}\neg\neg p$$

iff $p \in T$ for some $T \in \mathcal{T}$.

When for an EHT model $(\mathcal{T},\hbar)$ and a set of designated worlds $\mathcal{T}_0 \subseteq \mathcal{T}$ we have $(\mathcal{T},\hbar), T \models_{\mathsf{EHT}} \varphi$ for every $T \in \mathcal{T}_0$ then we write $(\mathcal{T},\hbar), \mathcal{T}_0 \models_{\mathsf{EHT}} \varphi$ for short. We call $((\mathcal{T},\hbar), \mathcal{T}_0)$ a *multipointed* EHT model of φ. Finally, we write $(\mathcal{T},\hbar), \mathcal{T}_0 \models_{\mathsf{EHT}} \Phi$ when $(\mathcal{T},\hbar), \mathcal{T}_0 \models_{\mathsf{EHT}} \varphi$ for every $\varphi \in \Phi$.

Proposition 3.1. *The following are equivalent.*

- $(\mathcal{T},\hbar), \mathcal{T} \models_{\mathsf{EHT}} \Phi$*;*
- $(\mathcal{T},\hbar), T \models_{\mathsf{EHT}} \mathsf{K}(\bigwedge \Phi)$ *for every* $T \in \mathcal{T}$*;*
- $(\mathcal{T},\hbar), T \models_{\mathsf{EHT}} \mathsf{K}(\bigwedge \Phi)$ *for some* $T \in \mathcal{T}$*.*

An EHT model $(\mathcal{T},\hbar)$ can alternatively be described as a collection $\{(\hbar(T), T)\}_{T\in\mathcal{T}}$ of HT models. We display concrete pointed and multipointed models as collections $\{(\hbar(T), T)\}_{T\in\mathcal{T}}$ where the actual worlds are underlined. Here are some examples:

$$\begin{array}{llrl}
\{(\emptyset,\{p\}), \underline{(\emptyset,\{q\})}\} \models_{\mathsf{EHT}} \neg p & & \{\underline{(\emptyset,\emptyset)}\} \models_{\mathsf{EHT}} & \neg\mathsf{K}\neg p \to p \\
\{\underline{(\emptyset,\{p\})}, (\emptyset,\{q\})\} \models_{\mathsf{EHT}} \neg\neg p & & \{\underline{(\{p\},\{p\})}\} \models_{\mathsf{EHT}} & \neg\mathsf{K}\neg p \to p \\
\{\underline{(\emptyset,\{p\})}, \underline{(\emptyset,\{q\})}\} \models_{\mathsf{EHT}} \neg\mathsf{K}\neg p & & \{\underline{(\emptyset,\{p\})}\} \not\models_{\mathsf{EHT}} & \neg\mathsf{K}\neg p \to p
\end{array}$$

Observe that the satisfaction of formulas of the form $\mathsf{K}\varphi$, $\neg\mathsf{K}\varphi$, $\neg\neg\mathsf{K}\varphi$, $\hat{\mathsf{K}}\varphi$, $\neg\hat{\mathsf{K}}\varphi$ and $\neg\neg\hat{\mathsf{K}}\varphi$ does not depend on the designated worlds: when $(\mathcal{T},\hbar), \mathcal{T}_0 \models_{\mathsf{EHT}} \mathsf{K}\varphi$ for *some* $\mathcal{T}_0 \subseteq \mathcal{T}$ then $(\mathcal{T},\hbar), \mathcal{T}_0 \models_{\mathsf{EHT}} \mathsf{K}\varphi$ for *every* $\mathcal{T}_0 \subseteq \mathcal{T}$, etc.

The next result is the heredity property in EHT: if a formula has an EHT model then it also has a classical $\mathsf{S5}$ model.

Proposition 3.2. *If* $(\mathcal{T},\hbar), T \models_{\mathsf{EHT}} \varphi$ *then* $(\mathcal{T}, id), T \models_{\mathsf{EHT}} \varphi$*.*

We now list some useful properties.

Proposition 3.3. *Let* φ *be an* EHT *formula. Then:*

1. $(\mathcal{T},\hbar), T \models_{\mathsf{EHT}} \neg\varphi$ *iff* $(\mathcal{T},id), T \not\models_{\mathsf{EHT}} \varphi$*;*
2. $(\mathcal{T},\hbar), T \models_{\mathsf{EHT}} \neg\neg\varphi$ *iff* $(\mathcal{T},id), T \models_{\mathsf{EHT}} \varphi$*;*
3. $(\mathcal{T},\hbar), T \models_{\mathsf{EHT}} \neg\mathsf{K}\varphi$ *iff* $(\mathcal{T},id), T' \not\models_{\mathsf{EHT}} \varphi$ *for some* $T' \in \mathcal{T}$*;*
4. $(\mathcal{T},\hbar), T \models_{\mathsf{EHT}} \neg\hat{\mathsf{K}}\varphi$ *iff* $(\mathcal{T},id), T' \not\models_{\mathsf{EHT}} \varphi$ *for every* $T' \in \mathcal{T}$*.*

We end with an appropriate definition of bisimilarity. Let $(\mathcal{T}_1, \hbar_1)$ and $(\mathcal{T}_2, \hbar_2)$ be two EHT models and $P \subseteq \mathbb{P}$. A relation $Z \subseteq \mathcal{T}_1 \times \mathcal{T}_2$ is called a *P-bisimulation* if

1. both Z and Z^{-1} are serial[4], and
2. if $T_1 Z T_2$ then $T_1 \cap P = T_2 \cap P$ and $\hbar_1(T_1) \cap P = \hbar_2(T_2) \cap P$.

If there exists a P-bisimulation Z between $(\mathcal{T}_1, \hbar_1)$ and $(\mathcal{T}_2, \hbar_2)$ such that $T_1 Z T_2$ then T_1 and T_2 are called *P-bisimilar*.

Proposition 3.4. *Let* φ *be an* EHT *formula. Let* $((\mathcal{T}_1, \hbar_1), T_1)$ *and* $((\mathcal{T}_2, \hbar_2), T_2)$ *be two pointed* EHT *models such that* T_1 *and* T_2 *are* $\mathbb{P}_\varphi$*-bisimilar. Then*

$$(\mathcal{T}_1, \hbar_1), T_1 \models_{\mathsf{EHT}} \varphi \quad \textit{iff} \quad (\mathcal{T}_2, \hbar_2), T_2 \models_{\mathsf{EHT}} \varphi.$$

3.4 EHT validity

A formula φ is called EHT *valid* if $(\mathcal{T},\hbar), \mathcal{T} \models_{\mathsf{EHT}} \varphi$ for every EHT model $(\mathcal{T},\hbar)$.

All principles of intuitionistic modal logics that were studied in the literature [Simpson, 1994] are EHT valid. As always in intuitionistic modal logics, K and $\hat{\mathsf{K}}$ are not dual: while $\hat{\mathsf{K}}\varphi \to \neg\mathsf{K}\neg\varphi$ is valid, the other direction is not. Here are some more examples: $\mathsf{K}\varphi \to \neg\hat{\mathsf{K}}\neg\varphi$, $\neg\hat{\mathsf{K}}\neg\varphi \to \neg\neg\hat{\mathsf{K}}\varphi$ and $\neg\neg\mathsf{K}\varphi \to \neg\mathsf{K}\neg\varphi$ are all valid while their converses are not. On the other hand, none of $\neg\neg\mathsf{K}\varphi \to \mathsf{K}\varphi$ and $\mathsf{K}\neg\neg\varphi \to \mathsf{K}\varphi$ is EHT valid. The same holds if we replace K by $\hat{\mathsf{K}}$. (The EHT model $\{(\emptyset, \{p\})\}$ provides a countermodel for all examples above.) However:

Proposition 3.5. *The equivalences* $\neg\mathsf{K}\varphi \leftrightarrow \hat{\mathsf{K}}\neg\varphi$ *and* $\neg\hat{\mathsf{K}}\varphi \leftrightarrow \mathsf{K}\neg\varphi$ *are* EHT *valid.*

4 Epistemic and autoepistemic equilibrium models

Pearce defined equilibrium models (EMs) of a formula as classical models satisfying a minimality condition when viewed as total HT models. Our epistemic equilibrium models (EEMs) generalise such models from classical to $\mathsf{S5}$ models in a natural way. Such models however only minimise truth, while the semantics of epistemic specifications involves

[4]The relation $Z \subseteq \mathcal{T}_1 \times \mathcal{T}_2$ is serial when for every $T_1 \in \mathcal{T}_1$ there is a $T_2 \in \mathcal{T}_2$ such that $(T_1, T_2) \in Z$.

formula	world views	EEMs
$\mathsf{K}\, p \to p$	$\{\emptyset\}$	$\{\emptyset\}$
$\neg\mathsf{K}\,\neg p \to p$	$\{\{p\}\}$	$\{\emptyset\}$ and $\{\{p\}\}$
$(p \vee q) \wedge (\neg\mathsf{K}\,\neg q \to p)$	$\{\{p\}\}$	$\{\{p\}\}$
$(p \vee q) \wedge (\mathsf{K}\, p \to r)$	$\{\{p\},\{q\}\}$	$\{\{p\},\{q\}\}$ and $\{\{p,r\}\}$

Table 3: The EHT counterparts of the epistemic specifications of Table 2, their world views and their EEMs.

also a minimisation of knowledge. We therefore define autoepistemic equilibrium models (AEEMs) as EEMs that are maximal under some order that comprises set inclusion.

4.1 Epistemic equilibrium models

We have already observed that total EHT models can be identified with classical S5 models: for $\mathcal{T} \subseteq 2^{\mathbb{P}}$ and $T_0 \in \mathcal{T}$, we write $\mathcal{T}, T_0 \models_{\mathsf{S5}} \varphi$ instead of $(\mathcal{T}, id), T_0 \models_{\mathsf{EHT}} \varphi$. We define the *epistemic equilibrium models* of φ as particular S5 models:

$$EEM(\varphi) = \{\mathcal{T} \subseteq 2^{\mathbb{P}} \;:\; \mathcal{T}, \mathcal{T} \models_{\mathsf{S5}} \varphi \text{ and there is no } h \neq id \text{ such that } (\mathcal{T}, \hbar), \mathcal{T} \models_{\mathsf{EHT}} \varphi\}$$

Table 3 illustrates our definition by means of the EHT counterparts of the epistemic specifications of Table 2 together with their world views and EEMs. Note that not every EEM is also a world view.

Proposition 4.1. *Let φ be an* EHT *formula.*

1. *If φ is* EHT *valid then $EEM(\varphi) = \big\{\{\emptyset\}\big\}$.*
2. *If $\{\emptyset\} \models_{\mathsf{S5}} \varphi$ then $EEM(\neg\neg\varphi) = \big\{\{\emptyset\}\big\}$ and $EEM(\neg\varphi) = \emptyset$.*
   If $\{\emptyset\} \not\models_{\mathsf{S5}} \varphi$ then $EEM(\neg\neg\varphi) = \emptyset$ and $EEM(\neg\varphi) = \big\{\{\emptyset\}\big\}$.
3. *$EEM(\mathsf{K}\,\varphi) = EEM(\varphi)$.*
4. *If $EEM(\varphi) = \emptyset$ then $EEM(\hat{\mathsf{K}}\,\varphi) = \emptyset$.*

Proposition 4.2. *Let φ be a non-modal* EHT *formula. Let $EM(\varphi)$ be the set of (classical) equilibrium models of φ.*

$$EEM(\varphi) = \{\mathcal{T} \subseteq EM(\varphi) \;:\; \mathcal{T} \neq \emptyset\}$$

$$EEM(\hat{\mathsf{K}}\,\varphi) = \begin{cases} \{\{T\} \;:\; T \in EM(\varphi)\} & \text{if } \emptyset \in EM(\varphi) \\ \{\{T\} \;:\; T \in EM(\varphi)\} \cup \{\{T, \emptyset\} \;:\; T \in EM(\varphi)\} & \text{otherwise} \end{cases}$$

We illustrate the above propositions by two examples. Remember that $EM(p \vee \neg p) = \{\emptyset, \{p\}\}$. It follows from Proposition 4.2 that $\hat{\mathsf{K}}\,(p \vee \neg p)$ has the EEMs $\{\emptyset\}$ and $\{\{p\}\}$ and that $p \vee \neg p$ has one more EEM, viz. $\{\emptyset, \{p\}\}$. It then follows from Proposition 4.1.3 that the EEMs of $\mathsf{K}\,(p \vee \neg p)$ are $\{\emptyset\}$, $\{\{p\}\}$ and $\{\emptyset, \{p\}\}$. Remember that $\neg\neg p$ has no EMs: the only candidate is $\emptyset$, but $\emptyset \not\models p$. So $\neg\neg p$, $\hat{\mathsf{K}}\,\neg\neg p$ and $\mathsf{K}\,\neg\neg p$ have no EEMs (cf. Proposition 4.1.2), and neither do $\neg\mathsf{K}\,\neg p$ and $\neg\hat{\mathsf{K}}\,\neg p$ (cf. Proposition 3.5).

4.2 Autoepistemic equilibrium models

While EEMs provide an interesting epistemic generalisation of EMs, they are somewhat too weak to provide an interesting semantics of epistemic specifications. Consider the 2nd epistemic specification $\neg\mathsf{K}\,\neg p \to p$ of Table 3: its EEM $\{\emptyset\}$ seems to contradict our intuitions about epistemic specifications. Roughly speaking, the interpretation of K in EEMs should quantify over all possible answer sets. We take that set to be $\bigcup EEM(\varphi)$ and we are going to select the most preferred EEMs in $EEM(\varphi)$ under set inclusion and an ordering that is determined by the answer sets from $\bigcup EEM(\varphi)$ that can be taken into account.

We start by defining a nonmonotonic satisfaction relation $\models^*$ for multipointed S5 models $(\mathcal{T}, \mathcal{T}_0)$ involving minimisation of truth over the set of designated worlds $\mathcal{T}_0$:

$$\mathcal{T}, \mathcal{T}_0 \models^* \varphi \text{ iff } \mathcal{T}, \mathcal{T}_0 \models_{\mathsf{S5}} \varphi \text{ and } (\mathcal{T}, \hbar), \mathcal{T}_0 \not\models_{\mathsf{EHT}} \varphi \text{ for every } \hbar \neq id \text{ that is total on } \mathcal{T} \setminus \mathcal{T}_0.$$

The first condition requires that for the total multipointed EHT model $(\mathcal{T}, id), \mathcal{T}_0$ we have $(\mathcal{T}, id), \mathcal{T}_0 \models_{\mathsf{EHT}} \varphi$. The second 'minimality of truth' condition, requires that there be no EHT model $((\mathcal{T}, \hbar), \mathcal{T}_0)$ of φ strictly weaker than $((\mathcal{T}, id), \mathcal{T}_0)$ on $\mathcal{T}_0$, where by the latter we understand that $\hbar(T) \subset T$ for some $T \in \mathcal{T}_0$ and $\hbar(T) = T$ for every $T \in \mathcal{T} \setminus \mathcal{T}_0$. Here are two examples (remember that the designated worlds are underlined):

- $\{\{p\}, \underline{\emptyset}\} \not\models^* \neg\mathsf{K}\,\neg p \to p$ because $\{\{p\}, \underline{\emptyset}\} \not\models_{\mathsf{S5}} \neg\mathsf{K}\,\neg p \to p$.
- $\{\underline{\{p\}}, \emptyset\} \models^* \neg\mathsf{K}\,\neg p \to p$ because $\{\underline{\{p\}}, \emptyset\} \models_{\mathsf{S5}} \neg\mathsf{K}\,\neg p \to p$ and $\{\underline{(\emptyset, \{p\})}, (\emptyset, \emptyset)\} \not\models_{\mathsf{EHT}} \neg\mathsf{K}\,\neg p \to p$.

Note that $EEM(\varphi) = \{\mathcal{T} \subseteq 2^{\mathbb{P}} \;:\; \mathcal{T}, \mathcal{T} \models^* \varphi\}$.

Let φ be an EHT formula. We define a *φ-indexed partial preorder* over S5 models by:

$$\mathcal{T} \leq_\varphi \mathcal{S} \text{ iff for every } T \in EEM(\varphi), \text{ if } \mathcal{T} \cup \{T\} \models^* \varphi \text{ then } \mathcal{S} \cup \{T\} \models^* \varphi.$$

The strict version of $\leq_\varphi$ is defined as usual:

$$\mathcal{T} <_\varphi \mathcal{S} \text{ iff } \mathcal{T} \leq_\varphi \mathcal{S} \text{ and } \mathcal{S} \not\leq_\varphi \mathcal{T}.$$

When $\mathcal{T} <_\varphi \mathcal{S}$ we say that $\mathcal{S}$ is *preferred* over $\mathcal{T}$ for φ.

We are then interested in EEMs of φ that are maximal w.r.t. set inclusion and w.r.t. $\leq_\varphi$: we say that $\mathcal{T}$ is an *autoepistemic equilibrium model* (AEEM) of φ if

1. $\mathcal{T} \in EEM(\varphi)$;
2. there is no $\mathcal{S} \in EEM(\varphi)$ such that $\mathcal{T} \subset \mathcal{S}$;
3. there is no $\mathcal{S} \in EEM(\varphi)$ such that $\mathcal{T} <_\varphi \mathcal{S}$.

For example, $\mathsf{K}\,(p \vee q)$ has three EEMs $\{\{p\}\}$, $\{\{q\}\}$ and $\{\{p\}, \{q\}\}$. While they are incomparable under $\leq_\varphi$, the latter is the only AEEM because it is set-inclusion maximal. Consider again $\Pi^* = \neg\mathsf{K}\,\neg p \to p$ of Table 3. For its two EEMs $\{\emptyset\}$ and $\{\{p\}\}$ we have $\{\emptyset\} <_{\Pi^*} \{\{p\}\}$. This is the case because $\{\emptyset, \underline{\{p\}}\} \models^* \Pi^*$ while $\{\underline{\emptyset}, \{p\}\} \not\models^* \Pi^*$. So $\{\{p\}\}$ is the only AEEM of Π^*, matching Kahl's semantics.

When φ has an EEM then φ also has an AEEM. It follows from Proposition 4.2 that a nonmodal formula φ has at most AEEM: the set of all EMs of φ.

Proposition 4.3. *If $EEM(\varphi) = \emptyset$ then φ has no AEEM. If $EEM(\varphi)$ is a singleton then then the AEEMs of φ equal the EEMs of φ.*

epistemic specif. Π	world views	EEMs of Π^*
p or $q \leftarrow$	$\{\{p\},\{q\}\}$	$\{\{p\}\} \subset \{\{p\},\{q\}\}$ $\{\{q\}\} \subset \{\{p\},\{q\}\}$
p or $q \leftarrow$ $p \leftarrow$ not K not q	$\{\{p\}\}$	$\{\{p\}\}$
p or $q \leftarrow$ $p \leftarrow \mathsf{K}\,q$	$\{\{p\},\{q\}\}$	$\{\{p\}\} \subset \{\{p\},\{q\}\}$
p or $q \leftarrow$ $p \leftarrow$ not K q	$\{\{p\}\}$	$\{\{q\}\} <_{\Pi^*} \{\{p\}\}$
p or $q \leftarrow$ $r \leftarrow \mathsf{K}\,p$	$\{\{p\},\{q\}\}$	$\{\{p,r\}\} <_{\Pi^*} \{\{p\},\{q\}\}$
p or $q \leftarrow$ $r \leftarrow$ not K not p	$\{\{p,r\},\{q,r\}\}$	$\{\{p,r\}\}\subset\{\{p,r\},\{q,r\}\}$ $\{\{q\}\}<_{\Pi^*}\{\{p,r\},\{q,r\}\}$
$p \leftarrow \mathsf{K}\,p$	$\{\emptyset\}$	$\{\emptyset\}$
$p \leftarrow$ not K not p	$\{\{p\}\}$	$\{\emptyset\} <_{\Pi^*} \{\{p\}\}$
$p \leftarrow \mathsf{K}\,p$ $p \leftarrow$ not K p	none	none
$p \leftarrow$ not K q $q \leftarrow$ not K p	$\{\{p\}\}$ and $\{\{q\}\}$	$\{\{p\}\}$ and $\{\{q\}\}$ (incomparable)
$p \leftarrow$ not K not p $p \leftarrow$ K not p	$\{\{p\}\}$	$\{\{p\}\}$
$p \leftarrow q$ $q \leftarrow$ not K not p	$\{\{p,q\}\}$	$\{\emptyset\} <_{\Pi^*} \{\{p,q\}\}$
$p \leftarrow$ not K not q $q \leftarrow$ not K not p	$\{\{p,q\}\}$	$\{\emptyset\} <_{\Pi^*} \{\{p,q\}\}$
$p \leftarrow$ not K not q, not q $q \leftarrow$ not K not p, not p	$\{\emptyset\}$ and $\{\{p\},\{q\}\}$	$\{\emptyset\} <_{\Pi^*} \{\{p\},\{q\}\}$

Table 4: Examples of Kahl's world views and our AEEMs, which are the maximal EEMs under $\subset$ or $\leq_{\Pi^*}$.

5 AEEMs for epistemic specifications

We now compare AEEMs with existing semantics for epistemic specifications. But first we translate epistemic specifications into EHT formulas.

5.1 Embedding epistemic specifications

Let Π be an epistemic specification. Our translation $(.)^*$ replaces '$\leftarrow$', '`or`', ',' and '`not`' respectively by '$\rightarrow$', '$\vee$', '$\wedge$' and '$\neg$'. Furthermore, it introduces a fresh variable $\tilde{p}$ for each $\sim p$ occurring in Π. For these new variables, the formula $\mathsf{Cons}(\Pi) = \bigwedge_{p \in \mathbb{P}_\Pi} \neg(p \wedge \tilde{p})$ ensures that p and $\tilde{p}$ cannot be true at the same time. (We only need it for those p that are prefixed by a strong negation in Π.) Here is an example:

$$\Pi = p \,\texttt{or} \sim q \leftarrow r, \texttt{not}\, s$$
$$\Pi^* = ((r \wedge \neg s) \rightarrow (p \vee \tilde{q})) \wedge \neg(q \wedge \tilde{q})$$

5.2 Comparison with Kahl's semantics

We now compare Kahl's world views of an epistemic specification Π with our AEEMs of Π^*. We do so by means of a series of examples most of which stem from [Kahl, 2014].

Table 4 lists epistemic specifications together with their (ordered) EEMs. The only example where the two semantics differ is the last one. The intuitions of quantification about answer sets seem to argue in favour of our semantics here.

Let us terminate with the AEEMs of Gelfond's 'eligibility' specification Π_G of Section 2.1. The formula Π_G^* has three EEMs: $\mathcal{T}_1 = \{\{h,e,i\},\{f,i\}\}$, $\mathcal{T}_2 = \{\{h,e\}\}$ and $\mathcal{T}_3 = \{\{f,i\}\}$ of which only $\mathcal{T}_1$ is intended. It can be checked that $\mathcal{T}_3 \subset \mathcal{T}_1$ and $\mathcal{T}_2 <_{\Pi_G^*} \mathcal{T}_1$.

5.3 Comparison with older approaches

We here discuss some older approaches that tried to generalise or to refine epistemic specifications and then compare our approach to an epistemic extension of equilibrium logic that is due to Wang and Zhang [2005] and that is close in spirit to our approach. All the approaches that we discuss here deal with the former version of epistemic specifications of [Gelfond, 1991; 1994; Baral and Gelfond, 1994], which behaves differently from the more recent versions in [Gelfond, 2011; Kahl, 2014].

As far as we know, the first embedding of epistemic specifications into an epistemic modal logic with a kind of minimal model reasoning about epistemic concepts knowledge and belief was Chen's [1997]. His approach significantly differs from ours because our approach is in terms of an intuitionistic modal logic with non-dual epistemic operators. More recently, Truszczyński [2011] proposed a refinement of epistemic specifications. His approach allows for the model $\{\emptyset\}$ of the formula $(p \vee \neg p) \rightarrow p$, which departs from all other approaches (in particular it has no EMs).

Wang and Zhang (WZ) [2005] described an epistemic extension of equilibrium logic into which they were able to embed Gelfond's first version of epistemic specifications. This extension also gives semantics to nested epistemic logic programs. The language extends that of HT by two modal operators K and M. Let us call WZ-EHT models their epistemic HT models. A WZ-EHT model is a triple $(\mathcal{A}, H, T)$ where $\mathcal{A} \subseteq 2^{\mathbb{P}}$ is a collection of valuations and (H,T) is an HT model. Note that H and T are not necessarily contained in $\mathcal{A}$. Given $\mathcal{A}$, define the collection

$$coll(\mathcal{A}) = \{(H,T) \ : \ H, T \in \mathcal{A} \text{ such that } H \subseteq T\}.$$

Then the definition of the satisfaction relation can be recast in terms of our semantics as follows:

$\mathcal{A}, H, T \models_{\mathsf{EHT}} p$ if $p \in H$;
$\mathcal{A}, H, T \models_{\mathsf{EHT}} \mathsf{K}\varphi$ if $coll(\mathcal{A}), A \models_{\mathsf{EHT}} \varphi$ for all $A \in coll(\mathcal{A})$;
$\mathcal{A}, H, T \models_{\mathsf{EHT}} \mathsf{M}\varphi$ if $coll(\mathcal{A}), A \models_{\mathsf{EHT}} \varphi$ for some $A \in coll(\mathcal{A})$.

Then WZ-EEMs for a theory Φ are total WZ-EHT models $(\mathcal{A}, T, T)$ of Φ such that there is no EHT model $(\mathcal{A}, H, T)$ of Φ with $H \subset T$. Finally, Wang and Zhang define *equilibrium views* of Φ to be the maximal collections $\mathcal{A} \subset 2^{\mathbb{P}}$ satisfying the fixed-point equation

$$\mathcal{A} = \{T \ : \ (\mathcal{A}, T, T) \text{ is a WZ-EEM of } \Phi\}.$$

While both we and Wang and Zhang propose an epistemic extension of equilibrium logic, there are some fundamental differences. Let us summarise the relationship. First, each satisfiable EHT formula has also a WZ-EHT model; the other way round, for instance $\mathsf{K}\,p \wedge \neg p$ has a WZ-EHT model but no EHT model. Second, the concept of minimality of knowledge differs: both $\mathsf{K}\,p \wedge \neg p$ and $\mathsf{K}\,\neg\neg p$ have WZ-EEMs but no EEMs in our sense; the other way round, $\mathsf{K}\,p \wedge \neg\neg p$ has an EEM in our sense, but no WZ-EEM. Third, maximisation of ignorance is not performed in the same way: $\mathsf{K}\,p$ has an AEEM, but no WZ-equilibrium view. The formulas $\hat{\mathsf{K}}\,p$, $\hat{\mathsf{K}}\,p \wedge \neg p$, $(p \vee q) \wedge (\hat{\mathsf{K}}\,q \rightarrow p)$ and $p \vee \mathsf{K}\,(p \vee \neg p)$ are other examples; $(\hat{\mathsf{K}}\,\neg p \vee \hat{\mathsf{K}}\,\neg\neg p) \wedge (\hat{\mathsf{K}}\,p \rightarrow (p \vee \neg p))$ has a WZ-equilibrium view, but no AEEM.

6 Strong equivalence

We now show that strong equivalence both in the EEM sense and in the AEEM sense can be captured in EHT. Precisely, we show that EHT validity of the equivalence

$$\mathsf{K}\big(\bigwedge \Phi_1\big) \leftrightarrow \mathsf{K}\big(\bigwedge \Phi_2\big)$$

characterises that

1. for every Θ, $\Phi_1 \cup \Theta$ and $\Phi_2 \cup \Theta$ have the same EEMs;
2. for every Θ, $\Phi_1 \cup \Theta$ and $\Phi_2 \cup \Theta$ have the same AEEMs.

Our proofs are non-trivial generalisations of Lifschitz et al.'s [2001] proof for HT logic. They require the syntactical characterisation of EHT models. To that end, using Proposition 3.4, we may suppose without loss of generality that $\mathbb{P}$ is *finite*. This allows to describe possible worlds and EHT models by means of formulas.

6.1 Some characteristic formulas

Let us define the following:

$$\mathsf{Total} = \mathsf{K} \bigwedge_{p\in\mathbb{P}} (p \vee \neg p)$$

$$\mathsf{At}(T) = \big(\bigwedge_{p\in T} p\big) \wedge \big(\bigwedge_{p\notin T} \neg p\big)$$

$$\mathsf{Ch}(\hbar) = \mathsf{K} \bigwedge_{T\in\mathcal{T}} \Big(\neg\neg\mathsf{At}(T) \to \Big(\big(\bigwedge_{p\in\hbar(T)} p\big) \wedge \bigwedge_{p\in(T\setminus\hbar(T))} (p \to \mathsf{Total})\Big)\Big)$$

where $T \in 2^{\mathbb{P}}$ and $(\mathcal{T},\hbar)$ is an EHT model. The formula Total captures that the model is total. $\mathsf{At}(T)$ says that we are at a there-world (and not at a here-world). Given an S5 model $\mathcal{T}$, $\mathsf{Ch}(\hbar)$ characterises $\hbar$: it says that the here-function is either $\hbar$ or *id*.

Lemma 6.1. *Let* $((\mathcal{T},\hbar),\mathcal{T}_0)$ *be a multipointed* EHT *model and let* $T_0, T \in \mathcal{T}$. *Then:*

1. $(\mathcal{T},\hbar), T_0 \models_{\mathsf{EHT}} \mathsf{Total}$ *iff* $\hbar = id$.
2. $(\mathcal{T},\hbar), T_0 \models_{\mathsf{EHT}} \mathsf{At}(T)$ *iff* $T_0 = T = \hbar(T_0)$.
3. $(\mathcal{T},\hbar), T_0 \models_{\mathsf{EHT}} \neg\neg\mathsf{At}(T)$ *iff* $T_0 = T$.
4. $(\mathcal{T},\hbar), \mathcal{T}_0 \models_{\mathsf{EHT}} \mathsf{Ch}(h)$ *and* $\mathcal{T}, \mathcal{T}_0 \models_{\mathsf{S5}} \mathsf{Ch}(h)$.
5. *If* $(\mathcal{T},\hbar'), \mathcal{T}_0 \models_{\mathsf{EHT}} \mathsf{Ch}(h)$ *then* $\hbar' = id$ *or* $\hbar' = \hbar$.

Proof. We only prove the last item.

Let $(\mathcal{T},\hbar'), \mathcal{T}_0 \models_{\mathsf{EHT}} \mathsf{Ch}(h)$. We use that by Lemma 6.1.3, $(\mathcal{T},\hbar'), T \models_{\mathsf{EHT}} \neg\neg\mathsf{At}(T)$. Suppose $\hbar' \neq \hbar$. Then there are $T \in \mathcal{T}$ and $p \in T$ such that either $p \in \hbar(T)$ and $p \notin \hbar'(T)$, or $p \notin \hbar(T)$ and $p \in \hbar'(T)$. In the former case, the hypothesis $(\mathcal{T},\hbar'), \mathcal{T}_0 \models_{\mathsf{EHT}} \mathsf{Ch}(h)$ implies that $(\mathcal{T},\hbar'), \mathcal{T}_0 \models_{\mathsf{EHT}} p$, i.e., a contradiction. In the latter case, as $p \notin \hbar(T)$ and $p \in T$ it implies that $(\mathcal{T},\hbar'), T \models_{\mathsf{EHT}} \mathsf{Total}$; then by Lemma 6.1.1 we must have $\hbar' = id$. □

6.2 Strong equivalence for EEMs

Lemma 6.2. *Let* $((\mathcal{T},\hbar),\mathcal{T}_0)$ *be a multipointed* EHT *model such that* $\hbar$ *is total on* $\mathcal{T} \setminus \mathcal{T}_0$. *Then* $(\mathcal{T},\hbar),\mathcal{T}_0 \models_{\mathsf{EHT}} \Phi_1$ *and* $(\mathcal{T},\hbar),\mathcal{T}_0 \not\models_{\mathsf{EHT}} \Phi_2$ *implies that there is a theory* Θ *such that*

(a) $\mathcal{T},\mathcal{T}_0 \not\models^* \Phi_1 \cup \Theta$ *and* $\mathcal{T},\mathcal{T}_0 \models^* \Phi_2 \cup \Theta$, *or*

(b) $\mathcal{T},\mathcal{T}_0 \models^* \Phi_1 \cup \Theta$ *and* $\mathcal{T},\mathcal{T}_0 \not\models^* \Phi_2 \cup \Theta$.

Proof. We build two different Θ, depending on whether $\mathcal{T},\mathcal{T}_0 \models_{\mathsf{S5}} \Phi_2$ or not.

When $\mathcal{T},\mathcal{T}_0 \not\models_{\mathsf{S5}} \Phi_2$ then we define $\Theta = \{\mathsf{Total}\}$ and prove that $\mathcal{T},\mathcal{T}_0 \models^* \Phi_1 \cup \Theta$ and $\mathcal{T},\mathcal{T}_0 \not\models^* \Phi_2 \cup \Theta$. The latter immediately follows from the hypothesis that $\mathcal{T},\mathcal{T}_0 \not\models_{\mathsf{S5}} \Phi_2$. As to the former, $\mathcal{T},\mathcal{T}_0 \models_{\mathsf{S5}} \Phi_1 \cup \Theta$ is the case because $\mathcal{T},\mathcal{T}_0 \models_{\mathsf{S5}} \Phi_1$ by hypothesis and heredity and because $\mathcal{T},\mathcal{T}_0 \models_{\mathsf{S5}} \mathsf{Total}$ by Lemma 6.1; moreover, for every $\hbar' \neq id$ we have $(\mathcal{T},\hbar'),\mathcal{T}_0 \not\models_{\mathsf{EHT}} \mathsf{Total}$, again by Lemma 6.1.

When $\mathcal{T},\mathcal{T}_0 \models_{\mathsf{S5}} \Phi_2$ then we define $\Theta = \{\mathsf{Ch}(\hbar)\}$ and prove that $\mathcal{T},\mathcal{T}_0 \not\models^* \Phi_1 \cup \Theta$ and $\mathcal{T},\mathcal{T}_0 \models^* \Phi_2 \cup \Theta$. The former holds because we are in a situation where $(\mathcal{T},\hbar),\mathcal{T}_0 \models_{\mathsf{EHT}} \Phi_1 \cup \Theta$ for a function $\hbar$ that is total on $\mathcal{T} \setminus \mathcal{T}_0$ and that is different from *id* (because $\mathcal{T},\mathcal{T}_0 \models_{\mathsf{S5}} \Phi_2$ and $(\mathcal{T},\hbar),\mathcal{T}_0 \not\models_{\mathsf{EHT}} \Phi_2$). As to the latter: first, we have $\mathcal{T},\mathcal{T}_0 \models_{\mathsf{S5}} \Phi_2$ by hypothesis and $\mathcal{T},\mathcal{T}_0 \models_{\mathsf{S5}} \mathsf{Ch}(\hbar)$ by Lemma 6.1; second, consider some $\hbar' \neq id$ that is total on $\mathcal{T} \setminus \mathcal{T}_0$. On the one hand, when $\hbar' = \hbar$ then by hypothesis $(\mathcal{T},\hbar'),\mathcal{T}_0 \not\models_{\mathsf{EHT}} \Phi_2$; on the other hand, when $\hbar' \neq \hbar$ then $(\mathcal{T},\hbar'),\mathcal{T}_0 \not\models_{\mathsf{EHT}} \mathsf{Ch}(\hbar)$ by Lemma 6.1. In both cases, $(\mathcal{T},\hbar'),\mathcal{T}_0 \not\models_{\mathsf{EHT}} \Phi_2 \cup \{\mathsf{Ch}(\hbar)\}$. □

Theorem 6.1. *Let* Φ_1 *and* Φ_2 *be two* EHT *theories. The following are equivalent:*

1. *For every* Θ, $\Phi_1 \cup \Theta$ *and* $\Phi_2 \cup \Theta$ *have the same EEMs;*
2. $\mathsf{K}(\bigwedge \Phi_1) \leftrightarrow \mathsf{K}(\bigwedge \Phi_2)$ *is* EHT *valid.*

Proof. For the right-to-left direction, suppose $\mathsf{K}(\bigwedge \Phi_1) \leftrightarrow \mathsf{K}(\bigwedge \Phi_2)$ is EHT valid, i.e., no EHT model allows to distinguish $\mathsf{K}(\bigwedge \Phi_1)$ and $\mathsf{K}(\bigwedge \Phi_2)$. Then $(\mathcal{T},\hbar),\mathcal{T} \models_{\mathsf{EHT}} \Phi_1$ iff $(\mathcal{T},\hbar),\mathcal{T} \models_{\mathsf{EHT}} \Phi_2$ by Proposition 3.1. Therefore for every theory Θ, $(\mathcal{T},\hbar),\mathcal{T} \models_{\mathsf{EHT}} \Phi_1 \cup \Theta$ iff $(\mathcal{T},\hbar),\mathcal{T} \models_{\mathsf{EHT}} \Phi_2 \cup \Theta$. It follows that $EEM(\Phi_1 \cup \Theta) = EEM(\Phi_2 \cup \Theta)$.

For the left-to-right direction, suppose the equivalence $\mathsf{K}(\bigwedge \Phi_1) \leftrightarrow \mathsf{K}(\bigwedge \Phi_2)$ is EHT invalid. Suppose w.l.o.g. that $(\mathcal{T},\hbar), T_0 \models_{\mathsf{EHT}} \mathsf{K}(\bigwedge \Phi_1)$ and $(\mathcal{T},\hbar), T_0 \not\models_{\mathsf{EHT}} \mathsf{K}(\bigwedge \Phi_2)$. Then $(\mathcal{T},\hbar),\mathcal{T} \models_{\mathsf{EHT}} \Phi_1$ and $(\mathcal{T},\hbar),\mathcal{T} \not\models_{\mathsf{EHT}} \Phi_2$ by Proposition 3.1. By Lemma 6.2 (which is applicable because $\hbar$ is trivially total on $\mathcal{T} \setminus \mathcal{T} = \emptyset$), we have that there must exist a Θ such that either $\mathcal{T},\mathcal{T} \not\models^* \Phi_1 \cup \Theta$ and $\mathcal{T},\mathcal{T} \models^* \Phi_2 \cup \Theta$, or $\mathcal{T},\mathcal{T} \models^* \Phi_1 \cup \Theta$ and $\mathcal{T},\mathcal{T} \not\models^* \Phi_2 \cup \Theta$. So $EEM(\Phi_1 \cup \Theta) \neq EEM(\Phi_2 \cup \Theta)$. □

6.3 Strong equivalence for AEEMs

We define a further formula characterising the set of all there-worlds of a EHT model $(\mathcal{T},\hbar)$ by means of Jankov-Fine-like formulas [Blackburn *et al.*, 2001]:

$$\mathsf{JF}(\mathcal{T}) = \big(\bigwedge_{T\in\mathcal{T}} \hat{\mathsf{K}}\,\neg\neg\mathsf{At}(T)\big) \wedge \mathsf{K}\big(\bigvee_{T\in\mathcal{T}} \neg\neg\mathsf{At}(T)\big).$$

Lemma 6.3. *Let* $(\mathcal{S},\hbar)$ *be an* EHT *model. Then*

$$(\mathcal{S},\hbar),\mathcal{S}_0 \models_{\mathsf{EHT}} \mathsf{JF}(\mathcal{T}) \quad \text{iff} \quad \mathcal{S} = \mathcal{T}.$$

Theorem 6.2. *Let Φ_1 and Φ_2 be two* EHT *theories. The following are equivalent:*

1. *For every Θ, $\Phi_1 \cup \Theta$ and $\Phi_2 \cup \Theta$ have the same AEEMs;*
2. $\mathsf{K}(\bigwedge \Phi_1) \leftrightarrow \mathsf{K}(\bigwedge \Phi_2)$ *is* EHT *valid.*

Proof. The proof of the right-to-left direction follows the lines of that of Theorem 6.1.

The proof of the left-to-right direction also follows the lines of that of Theorem 6.1, it is only the construction of the set Θ of Lemma 6.2 that has to be adapted: the Jankov-Fine formula $\mathsf{JF}(\mathcal{T})$ has to be conjoined with the Θs in the proof of Lemma 6.2. Lemma 6.3 guarantees that there is exactly one EEM of $\Phi_2 \cup \Theta$ (resp. $\Phi_1 \cup \Theta$). It then follows from Proposition 4.3 that there also is exactly one AEEM of $\Phi_2 \cup \Theta$ (resp. $\Phi_1 \cup \Theta$). □

7 Conclusion

In this paper, we have designed a monotonic, simple and neat intuitionistic modal logic EHT: the two modal operators K and $\hat{\mathsf{K}}$ that we add to the original language of HT are interpreted in models that are straightforward combinations of the Kripke models for S5 and HT. Based on this logic, we have first defined epistemic equilibrium models (EEMs) by means of a truth-minimising criterion, and then autoepistemic equilibrium models (AEEMs) by maximising ignorance. The latter models are powerful enough to provide a new logical semantics not only for epistemic specifications, but also for programs with arbitrary nestings of K and $\rightarrow$. We have provided a strong equivalence characterisation and have compared our semantics with the existing semantics for epistemic specifications.

We believe that EHT and the epistemic and autoepistemic equilibrium logics that we have built on it provide a good starting point for further extensions of answer-set programs by modal concepts. A first step could be the extension from single-agent epistemic logic to multi-agent epistemic logic.

References

[Baral and Gelfond, 1994] Chitta Baral and Michael Gelfond. Logic programming and knowledge representation. *The Journal of Logic Programming*, 19:73–148, 1994.

[Bierman and de Paiva, 2000] Gavin M. Bierman and Valeria de Paiva. On an intuitionistic modal logic. *Studia Logica*, 65(3):383–416, 2000.

[Blackburn *et al.*, 2001] Patrick Blackburn, Maarten de Rijke, and Yde Venema. *Modal Logic*. Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, 2001.

[Chen, 1997] Jianhua Chen. The generalized logic of only knowing (gol) that covers the notion of epistemic specifications. *Journal of Logic and Computation*, 7(2):159–174, 1997.

[Fariñas del Cerro and Raggio, 1983] Luis Fariñas del Cerro and Andrés R. Raggio. Some results in intuitionistic modal logic. *Logique et Analyse Louvain*, 26(102):219–224, 1983.

[Fischer-Servi, 1976] Gisèle Fischer-Servi. On modal logic with an intuitionistic base. *Studia Logica*, 36(4):141–149, 1976.

[Gelfond, 1991] Michael Gelfond. Strong introspection. In Thomas L. Dean and Kathleen McKeown, editors, *AAAI*, pages 386–391. AAAI Press / The MIT Press, 1991.

[Gelfond, 1994] Michael Gelfond. Logic programming and reasoning with incomplete information. *Annals of Mathematics and Artificial Intelligence*, 12(1-2):89–116, 1994.

[Gelfond, 2011] Michael Gelfond. New semantics for epistemic specifications. In James P. Delgrande and Wolfgang Faber, editors, *Logic Programming and Nonmonotonic Reasoning - 11th International Conference, LPNMR 2011, Vancouver, Canada, May 16-19, 2011. Proceedings*, volume 6645 of *Lecture Notes in Computer Science*, pages 260–265. Springer, 2011.

[Kahl, 2014] Patrick Thor Kahl. *Refining the semantics for epistemic logic programs*. Phd thesis, Texas Tech University, May 2014.

[Lifschitz *et al.*, 1999] Vladimir Lifschitz, Lappoon R. Tang, and Hudson Turner. Nested expressions in logic programs. *Annals of Mathematics and Artificial Intelligence*, 25(3-4):369–389, 1999.

[Lifschitz *et al.*, 2001] Vladimir Lifschitz, David Pearce, and Agustín Valverde. Strongly equivalent logic programs. *ACM Transactions on Computational Logic*, 2(4):526–541, 2001.

[Pearce, 1996] David Pearce. A new logical characterisation of stable models and answer sets. In Jürgen Dix, Luís Moniz Pereira, and Teodor C. Przymusinski, editors, *NMELP*, volume 1216 of *Lecture Notes in Computer Science*, pages 57–70. Springer Verlag, 1996.

[Pearce, 2006] David Pearce. Equilibrium logic. *Annals of Mathematics and Artificial Intelligence*, 47(1-2):3–41, 2006.

[Simpson, 1994] Alex K. Simpson. *The proof theory and semantics of intuitionistic modal logic*. Phd thesis, University of Edinburgh, University of Edinburgh, College of Science and Engineering, School of Informatics, November 1994.

[Truszczyński, 2011] Mirosław Truszczyński. Revisiting epistemic specifications. In *Logic programming, knowledge representation, and nonmonotonic reasoning*, pages 315–333. Springer, 2011.

[Wang and Zhang, 2005] Kewen Wang and Yan Zhang. Nested epistemic logic programs. In Chitta Baral, Gianluigi Greco, Nicola Leone, and Giorgio Terracina, editors, *Logic Programming and Nonmonotonic Reasoning, 8th International Conference, LPNMR 2005, Diamante, Italy, September 5-8, 2005, Proceedings*, volume 3662 of *Lecture Notes in Computer Science*, pages 279–290. Springer, 2005.