



**HAL**  
open science

## Autoepistemic equilibrium logic and epistemic specifications

Ezgi Iraz Su, Luis Fariñas del Cerro, Andreas Herzig

► **To cite this version:**

Ezgi Iraz Su, Luis Fariñas del Cerro, Andreas Herzig. Autoepistemic equilibrium logic and epistemic specifications. *Artificial Intelligence*, 2020, 282, pp.103249. 10.1016/j.artint.2020.103249 . hal-02945872

**HAL Id: hal-02945872**

**<https://hal.science/hal-02945872>**

Submitted on 22 Sep 2020

**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: <https://oatao.univ-toulouse.fr/26365>

### Official URL:

<https://doi.org/10.1016/j.artint.2020.103249>

### To cite this version:

Su, Ezgi Iraz  and Fariñas del Cerro, Luis  and Herzig, Andreas   
*Autoepistemic equilibrium logic and epistemic specifications.* (2020) *Artificial Intelligence*, 282. 103249. ISSN 0004-3702.

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

# Autoepistemic equilibrium logic and epistemic specifications

Ezgi Iraz Su\*, Luis Fariñas del Cerro, Andreas Herzig

IRIT, CNRS, Université Paul Sabatier, 118 Route de Narbonne, F-31062 Toulouse Cedex 9, France

## A B S T R A C T

Epistemic specifications extend disjunctive answer-set programs by an epistemic modal operator that may occur in the body of rules. Their semantics is in terms of world views, which are sets of answer sets, and the idea is that the epistemic modal operator quantifies over these answer sets. Several such semantics were proposed in the literature. We here propose a new semantics that is based on the logic of here-and-there: we add epistemic modal operators to its language and define epistemic here-and-there models. We then successively define epistemic equilibrium models and autoepistemic equilibrium models. The former are obtained from epistemic here-and-there models in exactly the same way as Pearce's equilibrium models are obtained from here-and-there models, viz. by minimising truth; they provide an epistemic extension of equilibrium logic. The latter are obtained from the former by maximising the set of epistemic possibilities, and they provide a new semantics for Gelfond's epistemic specifications. For both semantics we establish a strong equivalence result: we characterise strong equivalence of two epistemic programs by means of logical equivalence in epistemic here-and-there logic. We finally compare our approach to the existing semantics of epistemic specifications and discuss which formalisms provide more intuitive results by pointing out some formal properties a semantics proposal should satisfy.

### Keywords:

Answer set programming  
Epistemic specifications  
Epistemic logic programs  
Here-and-there logic  
Equilibrium logic  
Modal logic

## 1. Introduction

Answer-set programming (ASP) is a successful logic-based problem solving approach in knowledge representation and reasoning [1]. The interpretation of an ASP program is given in terms of *answer sets*, which are classical models of the program that satisfy some minimality criterion. The negation of a propositional variable  $p$  is interpreted as non-membership in the answer set under concern. Gelfond was the first to criticise this as unsatisfactory, exhibiting an example where there are several answer sets and where one wants to check non-membership of  $p$  in *all* answer sets. In order to be able to quantify over answer sets, Gelfond extended the language of disjunctive logic programs [2] by allowing epistemic operators in rule bodies [3]. Such operators enable us to reason about incomplete information, understood as situations in which there are multiple answer sets of a program. He called programs with epistemic operators in rule bodies *epistemic specifications*. Together with co-authors, he proposed several semantics for epistemic specifications [4–7].

The semantics of epistemic specifications is in terms of *world views*, which are in structure collections of answer sets. In modal logic, this corresponds to S5 models, which are collections of valuations. Similar to answer set semantics, a world view  $\mathcal{S}$  of an epistemic specification  $\Pi$  is defined by means of the reduct  $\Pi^{\mathcal{S}}$  of  $\Pi$  with respect to  $\mathcal{S}$ . Basically

---

\* Corresponding author.

E-mail addresses: [ezgiraz@gmail.com](mailto:ezgiraz@gmail.com) (E. Iraz Su), [farinas@irit.fr](mailto:farinas@irit.fr) (L. Fariñas del Cerro), [herzig@irit.fr](mailto:herzig@irit.fr) (A. Herzig).

the construction is in two steps: first, we compute the reduct  $\Pi^S$ , eliminating the epistemic modal operators; second, we compute the set of all answer sets of  $\Pi^S$  and check whether it equals  $\mathcal{S}$ . If this is the case, then  $\mathcal{S}$  is said to be a world view of  $\Pi$ .

Subsequently several different semantics were proposed for epistemic specifications [8–10,5,6,11,12,7,13,14]. Some propose different definitions of reducts or world views, and some others propose semantics inspired by the Kripke semantics of modal logics. In this work, we introduce a new semantics that is based on an epistemic extension of Pearce’s equilibrium logic [15,16]. This had already been undertaken by Wang and Zhang [9], however for the somewhat outdated first version of Gelfond’s epistemic specifications [3,4]. Our version is closer to Gelfond’s [5] and Kahl’s [6] more recent versions.

Pearce’s equilibrium logic characterises *answer set* semantics of nonepistemic logic programs. The underlying here-and-there logic (HT) [17] characterises their *strong equivalence* [18]. We add two epistemic operators  $K$  and  $\hat{K}$  to the language of HT and define epistemic HT models, abbreviated EHT models. Then, we propose epistemic equilibrium models (EEMs) by generalising the usual minimality criterion, defined over HT models, to EHT models. Finally, we introduce autoepistemic equilibrium models (AEEMs) as EEMs that are maximal under set inclusion or under a preference ordering on EEMs. AEEMs provide a new logical semantics for epistemic specifications and even for more general nested epistemic logic programs. Moreover, we establish a strong equivalence result: we show that strong equivalence of two epistemic logic programs can be characterised as their logical equivalence in EHT.

Several researchers tried to refine [5,6,11,19,7,12,20,14] or generalise [8–10] the semantics of epistemic specifications. Some are based on the somewhat outdated version of the formalism [3,4], so they are out of our consideration. Kahl et al. [7] and Shen et al. [12] refine Gelfond’s somewhat corrected version [5]. We compare our semantics with these recent approaches. We demonstrate by means of examples that all other semantics differ from our approach and that Kahl et al.’s comes closest. We argue that ours is more interesting because it is mathematically elegant and provides more intuitive results for our list of examples.

The rest of the paper is organised as follows. Section 3 recalls epistemic specifications. Section 4 introduces epistemic here-and-there logic. Section 5 defines epistemic equilibrium models (EEMs) and autoepistemic equilibrium models (AEEMs). Section 6 compares AEEMs with the existing semantics of epistemic specifications, in particular with Kahl et al.’s and Shen et al.’s more recent versions, as well as older related work. Section 7 provides strong equivalence characterisations in terms of EHT equivalence for both kinds of models. Section 8 presents an alternative relational semantics for EHT, which is equivalent to the original functional semantics. Section 9 discusses some formal principles of epistemic specifications recently proposed by Cabalar et al., in particular and mainly, epistemic splitting property. Section 10 concludes with some final remarks and discussions about future work.<sup>1</sup>

## 2. The need for epistemic specifications in knowledge representation and artificial intelligence

As first recognised by Gelfond, with the addition of epistemic modalities, the language of an epistemic extension of ASP is better suited for reasoning out incomplete information. To illustrate this, let us attempt to formalise the *presumption of legal innocence*: “a person is legally innocent unless proven guilty”. One may attempt to formalise this by the following ASP statement

$$\text{innocent}(X) \leftarrow \text{not guilty}(X) \tag{1}$$

This formalisation works if it is considered with a complete list of guilty people. However, if the rule (1) is used together with the following incomplete statement

$$\text{guilty}(\text{john}) \text{ or } \text{guilty}(\text{bob}) \tag{2}$$

which narrows the guilt to two suspects, the rule (1) fails to prove John’s legal innocence. This is, of course, the expected result because the fact (2) has two answer sets:  $\{\text{guilty}(\text{john})\}$  and  $\{\text{guilty}(\text{bob})\}$ . When we consider (1) and (2) together, the resulting program has the following answer sets:  $\{\text{guilty}(\text{john}), \text{innocent}(\text{bob})\}$  and  $\{\text{guilty}(\text{bob}), \text{innocent}(\text{john})\}$ . Thus, a more accurate expression of the principle should say something like “John is assumed innocent if there is at least one answer set of the program not containing  $\text{guilty}(\text{john})$ ”. This can be formalised by using modal operators as follows:

$$\text{innocent}(X) \leftarrow \text{not } K \text{ guilty}(X) \tag{3}$$

The program composed of (2) and (3) has a unique world view:

$$\{ \{ \text{guilty}(\text{john}), \text{innocent}(\text{john}), \text{innocent}(\text{bob}) \}, \\ \{ \text{guilty}(\text{bob}), \text{innocent}(\text{john}), \text{innocent}(\text{bob}) \} \}.$$

<sup>1</sup> This is an extended version of a paper that was presented at IJCAI 2015 [19] and of a part of the Ph.D. thesis of one of the authors [21].

Clearly, the ability to accurately formalise this and other similar legal principles is important for building systems capable of legal reasoning. A similar construct, expressed by the rule

$$\text{alarm} \leftarrow \text{not } K\text{safe}$$

can be used in various security situations. For instance, existence of at least one answer set in which safety is not established will be a sufficient cause for alarm.

Another important application of epistemic specifications is the *Closed World Assumption* (CWA) saying that “*p* is assumed to be false if there is no evidence to the contrary”. It is expressed in ASP by  $\sim p \leftarrow \text{not } p$  where ‘ $\sim$ ’ is strong negation. However, this formalisation is problematic: take  $\Pi = \{p \text{ or } q, \sim p \leftarrow \text{not } p\}$ .  $\Pi$  has two answer sets  $\{p\}$  and  $\{q, \sim p\}$ . Thus, its answer to both queries  $p?$  and  $q?$  is “unknown” since none of them is included in all answer sets. This result is unintended. As already argued in [12], the CWA is formalised more adequately in epistemic specifications as  $\sim p \leftarrow \text{not } Kp$  (in the propositional case).

Another close example which can be more satisfactorily expressed in the language of epistemic specifications is Murphy’s law [12], saying that “*if something can go wrong and we cannot prove that it will not go wrong, then it will go wrong*”. This law can be formalised in the language of epistemic specifications by  $\text{gowrong}(X) \leftarrow \text{not } K\sim\text{gowrong}(X)$ .

There are several domains of artificial intelligence (AI), such as planning, cryptographic protocols, autonomous robot control, security, etc. in which the language of an epistemic extension of equilibrium logic (in particular, epistemic equilibrium logic) can represent aspects that are important for their formalisations. Moreover, it should actually be relevant for any practical application of ASP where additional expressive power beyond default negation is required to correctly represent incomplete information.

In *conformant planning*, the agents do not have exact knowledge in which state the system is, and they have no means of observation to learn it. However, the language of epistemic equilibrium logic offers the resources to formalise this type of uncertainty. For example, the negation of the knowledge operator called *epistemic negation* ensures that a certain proposition will not be true in all stable extensions. Another important mechanism is *integrity constraints* on the set of stable extensions: they allow us to eliminate situations that are considered to be impossible or false in each extension. In particular, Kahl and Leclerc [14] have developed an epistemic specification system accounting for integrity constraints and modelling conformant planning problems. Zhang and Zhang [22] also have shown that conformant planning problems can be naturally represented by epistemic specifications under a specific semantics they proposed. Cabalar et al. illustrate the intuition behind the use of some properties of epistemic specifications like epistemic splitting in conformant planning through a simple example (see Example 2 in [23]).

The formalisation of *cryptographic protocols* can also take advantage of the expressive power of epistemic equilibrium logic, given that it allows us to distinguish what may be true in some stable models from what is true in all stable models, and also allows us to extend what is considered implicitly false in the presence of epistemic operators. It is clear that the formalisation of cryptographic protocols needs to consider the set of all possible situations. In the framework of ASP, it means being able to reason about the set of stable extensions. An interesting approach of formalising cryptographic protocols using ASP has been carried out by Aiello and Massacci [24–26] in which attacks are simulated by the construction of specific plans. All these illustrate the need to simulate epistemic concepts to be able to directly constrain the agents’ knowledge, in other words epistemic extensions of ASP.

### 3. Epistemic specifications

Various versions of epistemic specifications were introduced by Gelfond [3–5]. Later, Kahl et al. [6,11,7] proposed a further improvement of Gelfond’s most recent version [5] of the formalism. Finally, Shen et al. [12,20] came up with a different approach that is based on a non-standard ASP semantics. However, it seems that a fully satisfactory semantics has still not been given. We here recall Kahl et al.’s most recent version [7], to which our approach is closest; our semantics agrees with this version of epistemic specifications on most examples that are given in their paper. We will argue in Section 6 that it is more intuitive for the examples where they differ.

#### 3.1. The language of epistemic specifications

The language of epistemic specifications extends that of disjunctive logic programming by the modal operator  $K$ . The formula  $K\varphi$  is read “ $\varphi$  is known to be true”.<sup>2</sup> Literals of this language are of three different kinds: *objective literals* ( $l$ ), *extended objective literals* ( $\lambda$ ) and *extended subjective literals* ( $L$ ). They are defined by the following grammar:

<sup>2</sup> The original presentation has also another modal operator  $M$  which is dual to  $K$ : instead of  $K\text{not } l$  it contains  $\text{not } Ml$ , and instead of  $\text{not } K\text{not } l$  it includes  $Ml$ . Due to this definability via  $K$ , we give the language of epistemic specifications without  $M$ . Moreover,  $M\varphi$  is (somewhat nonstandardly) read “ $\varphi$  may be believed to be true”, while “ $\varphi$  is compatible with the agent’s belief” would be the more standard reading of a modal operator that is dual to the knowledge operator  $K$ . We observe that our operator  $\hat{K}$  is slightly different from Gelfond’s  $M$ :  $K$  and  $\hat{K}$  are not dual while  $K$  and  $M$  are dual, that is,  $M$  is equivalent to  $\text{not } K\text{not}$ . We argue that this is an advantage of our approach, because epistemic HT logic is a particular intuitionistic logic and as usual in intuitionistic logics, duality of necessity and possibility should fail. Moreover, from the perspective of knowledge representation and reasoning, this property is a strength of our approach because it makes our logic more expressive compared to epistemic specifications.

**Table 1**  
Kahl's definition of reduct of a program w.r.t. a world view  $S$  [6,7].

Literal $L$	If $S \models L$	If $S \not\models L$
$K\lambda$	replace by $\lambda$	replace by $\perp$
$\text{not } K\lambda$	replace by $\top$	replace by $\text{not } \lambda$

$$l := p \mid \sim p$$

$$\lambda := l \mid \text{not } l$$

$$L := K\lambda \mid \text{not } K\lambda$$

in which  $p$  ranges over the set  $\mathbb{P}$  of propositional variables. We suppose that  $\mathbb{P}$  contains  $\top$  and  $\perp$  as well. However, we call a literal *basic* if it is different from both  $\top$  and  $\perp$ .

So the language has two negations: *strong negation* ' $\sim\varphi$ ' and *default negation* (alias *negation as failure*) ' $\text{not } \varphi$ '. The latter is read as " $\varphi$  is false by default" and expresses that when we have no acceptable support that provides a justification for  $\varphi$  in an answer set, we assume  $\varphi$  to be false in that answer set.<sup>3</sup> The term "strong" signals that ' $\sim\varphi$ ' implies ' $\text{not } \varphi$ ', but not the other way around.

A rule  $\rho$  is of the form  $\text{head}(\rho) \leftarrow \text{body}(\rho)$ :

$$l_1 \text{ or } \dots \text{ or } l_m \leftarrow G_1, \dots, G_n$$

in which the literals of the head ' $l_1 \text{ or } \dots \text{ or } l_m$ ' are objective literals and the literals of the body ' $G_1, \dots, G_n$ ' are extended literals (either extended objective literals or extended subjective literals). Note that  $\perp$  and  $\top$  can also appear as a conjunct of body and this is necessary to be able to give a coherent definition of reduct (see Table 1) where we replace some of body literals with  $\top$  or  $\perp$ . Moreover, we consider the head of  $\rho$  to be  $\perp$  ('false') if  $m = 0$  and the body of  $\rho$  to be  $\top$  ('true') if  $n = 0$ .

An *epistemic specification* is a finite set of rules. When a program  $\Pi$  does not contain  $K$ , we call it *nonepistemic* or *nonmodal*. Below is Gelfond's well-known 'eligibility' example [3] that was taken up in all papers on epistemic specifications:

$$\begin{aligned} \Pi_G = \{ & h \text{ or } f \leftarrow \\ & e \leftarrow h \\ & e \leftarrow f, m \\ & \sim e \leftarrow \sim h, \sim f \\ & i \leftarrow \text{not } Ke, \text{not } K\sim e \} \end{aligned}$$

in which  $h$  stands for "high GPA",  $f$  stands for "fair GPA",  $m$  stands for "minority",  $e$  stands for "eligible for scholarship", and  $i$  stands for "to be interviewed". Throughout this paper, we call the above-mentioned epistemic specification  $\Pi_G$ . The last rule reads: "If it is not known whether the student is eligible for scholarship then there should be an interview".

### 3.2. The semantics of epistemic specifications

Let  $S$  be a consistent set of basic objective literals, that is, a set of basic objective literals in which  $\sim p$  and  $p$  cannot occur together for every  $p \in \mathbb{P}$ . We call such sets *belief sets*. Satisfaction of extended objective literals in a belief set  $S$  is defined by:

$$\begin{aligned} S \models_{\text{ES}} l & \quad \text{if } l \in S \text{ or } l = \top; \\ S \models_{\text{ES}} \text{not } l & \quad \text{if } l \notin S \text{ or } l = \perp. \end{aligned}$$

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

$$\begin{aligned} \mathcal{S} \models_{\text{ES}} K\lambda & \quad \text{if } S \models_{\text{ES}} \lambda \text{ for every } S \in \mathcal{S}; \\ \mathcal{S} \models_{\text{ES}} \text{not } K\lambda & \quad \text{if } \mathcal{S} \not\models_{\text{ES}} K\lambda. \end{aligned}$$

For example,  $\{\{p\}\} \models_{\text{ES}} \text{not } K\text{not } p$  because  $\{\{p\}\} \not\models_{\text{ES}} K\text{not } p$ . The latter is the case because there is an element of  $\{\{p\}\}$ , viz.  $\{p\}$ , such that  $\{p\} \not\models_{\text{ES}} \text{not } p$ .

<sup>3</sup> Gelfond's original language is slightly different, with  $\sim KI$  instead of  $\text{not } KI$  and  $\sim MI$  instead of  $\text{not } MI$ . However, they have the same semantics, *mutatis mutandis*. Probably, Gelfond considered  $\sim KI$  because  $\text{not}$  performs locally, i.e., on each answer set separately. So, perhaps he thought that  $\text{not}$  should not precede a modal operator, providing quantification over answer sets and that strong negation  $\sim$  would suit better since it refers to explicit negation. However, Gelfond never explains the reason of this choice in his papers.

**Table 2**

Examples of epistemic specifications and their world views.

Epistemic specification	World views
$\{p \leftarrow Kp\}$	$\{\emptyset\}$
$\{p \leftarrow \text{not} K \text{not} p\}$	$\{\{p\}\}$
$\{p \text{ or } q \leftarrow\}$	$\{\{p\}, \{q\}\}$
$\{p \text{ or } q \leftarrow, p \leftarrow \text{not} K \text{not} q\}$	$\{\{p\}\}$
$\{p \text{ or } q \leftarrow, r \leftarrow Kp\}$	$\{\{p\}, \{q\}\}$

Let  $\Pi$  be an epistemic specification, and let  $\mathcal{S}$  be a non-empty collection of consistent sets of basic objective literals. Whether  $\mathcal{S}$  is a world view of  $\Pi$  is decided by computing the reduct of  $\Pi$  with respect to  $\mathcal{S}$  first, finding the set of all answer sets of this reduct, then checking a fixed-point equation, and finally by following a maximality condition among all models satisfying this fixed-point equation. The reduct  $\rho^{\mathcal{S}}$  of a rule  $\rho \in \Pi$  with respect to  $\mathcal{S}$  is obtained from  $\rho$  by eliminating the  $K$  operator according to Table 1. For instance, let  $\rho_G = i \leftarrow \text{not} K e, \text{not} K \sim e$ . Its reduct by  $\{\{h, e, i\}\}$  is  $\rho_G^{\{\{h, e, i\}\}} = i \leftarrow \text{not} e, \top$  and its reduct by  $\{\{h, e, i\}, \{f, i\}\}$  is  $\rho_G^{\{\{h, e, i\}, \{f, i\}\}} = i \leftarrow \top, \top$ . The reduct of  $\Pi$  with respect to  $\mathcal{S}$ , denoted by  $\Pi^{\mathcal{S}}$ , is  $\Pi^{\mathcal{S}} = \{\rho^{\mathcal{S}} : \rho \in \Pi\}$ . Clearly, when no epistemic operators occur in  $\Pi$  then  $\Pi^{\mathcal{S}} = \Pi$ , for every  $\Pi$ .

In order to define world views one should also define the set of epistemic negations:

$$\text{Ep}(\Pi) = \{\text{not} K \lambda : K \lambda \text{ or } \text{not} K \lambda \text{ occurs in } \Pi \text{ and } \lambda \text{ is an extended objective literal}\}.$$

To illustrate this set, consider the program  $\Gamma = \{t \leftarrow Kp, \text{not} K \text{not} q, \text{not} K r, K \text{not} s\}$ . Thus,  $\text{Ep}(\Gamma) = \{\text{not} K p, \text{not} K \text{not} q, \text{not} K r, \text{not} K \text{not} s\}$ . Next, we define the subset  $\Phi_{\mathcal{S}} = \{L \in \text{Ep}(\Pi) : \mathcal{S} \models_{\text{ES}} L\}$  with respect to a candidate world view  $\mathcal{S}$ . Finally,  $\mathcal{S}$  is a world view of  $\Pi$  if:

$$\mathcal{S} \stackrel{\text{fp}}{=} \text{AS}(\Pi^{\mathcal{S}}), \text{ and there is no } \mathcal{S}' \text{ such that } \mathcal{S}' \stackrel{\text{fp}}{=} \text{AS}(\Pi^{\mathcal{S}'}) \text{ and } \Phi_{\mathcal{S}'} \supset \Phi_{\mathcal{S}},$$

where  $\text{AS}(\Pi)$  denotes the set of all answer sets of  $\Pi$ , and  $\stackrel{\text{fp}}{=}$  refers to the *fixed point* equation. As the case ' $\text{not} K \text{not} l$ ' may introduce double negation ' $\text{not} \text{not} l$ ', for the computation of answer sets of such programs one has to resort to answer set programming with nested expressions [27] where  $\text{not} \text{not} l$  is not equivalent to  $l$ . Intuitively, world views collect answer sets while maximising ignorance. Observe that and it has at most one world view: the collection of all answer sets of  $\Pi$ . Therefore, the unique world view of nonmodal programs cannot contain two different belief sets  $S_1$  and  $S_2$  such that  $S_1 \subset S_2$  because a disjunctive logic program without negation as failure in the head parts of its rules or a double negation in the body parts cannot have such two answer sets [2, Lemma 1]. It follows that world views of nonepistemic programs containing  $\emptyset$  cannot contain any other belief sets either.

**Example 1.** The only world view of Gelfond's 'eligibility' program  $\Pi_G$  of Section 3.1 is  $\{\{h, e, i\}, \{f, i\}\}$ .

**Example 2.** The reduct of an epistemic specification  $\{p \leftarrow Kp\}$  by  $\{\emptyset\}$  is  $\{p \leftarrow \perp\}$  and its reduct by  $\{\{p\}\}$  is  $\{p \leftarrow p\}$ . Both reducts have exactly one answer set:  $\emptyset$ . However, only the former satisfies the fixed point equation  $\stackrel{\text{fp}}{=}$ . So, the unique world view is  $\{\emptyset\}$ .

**Example 3.** The reduct of an epistemic specification  $\{p \text{ or } q \leftarrow, p \leftarrow \text{not} K \text{not} q\}$  by  $\{\{p\}\}$  is  $\{p \text{ or } q \leftarrow, p \leftarrow \text{not} \text{not} q\}$  and its reduct by  $\{\{q\}\}$  is  $\{p \text{ or } q \leftarrow, p \leftarrow \top\}$ . Both have exactly one answer set:  $\{p\}$ ; therefore the former  $\{\{p\}\}$  is the unique world view since it satisfies the fixed point equation  $\stackrel{\text{fp}}{=}$ .

Table 2 contains more examples.

An epistemic specification  $\Pi$  is *consistent* if it has at least one world view; otherwise it is *inconsistent*.

Gelfond's definition of reduct [5] differs from Table 1 for the case  $\text{not} K \lambda$ : when  $\mathcal{S} \not\models \text{not} K \lambda$  then it is replaced by  $\perp$ . Due to this subtle difference, Gelfond gets the world views  $\{\emptyset\}$  and  $\{\{p\}\}$  for the second line of Table 2, and he gets no world view at all for Example 3. As Kahl argues in his papers, this is not as intuitive as what is obtained with his own definition of reduct. We agree with him and therefore present his semantics here.

While Kahl has eliminated the unsupported world views obtained in Gelfond's version, some unintended results with his semantics still remain, as we will discuss in Section 6.2.

#### 4. Epistemic here-and-there logic (EHT)

The logic of here-and-there (HT) is a three-valued monotonic 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  $H \subseteq T$ . We call  $H$  ‘here-valuation’ and  $T$  ‘there-valuation’. Pearce was the first to realise that this logic provides a good logical basis for answer set programming [15]. Its importance increased even more when Lifschitz et al. proved that strong equivalence of logic programs can be characterised in HT logic [18].

In this section we introduce epistemic HT logic (EHT), which extends HT by two epistemic modal operators  $K$  and  $\hat{K}$  in the spirit of intuitionistic modal logics [28–31] (where the duality of modal operators  $K$  and  $\hat{K}$  fails). Basically, epistemic HT models (EHT models) generalise HT models  $(H, T)$  to collections  $\{(H_i, T_i)\}_i$  of such models. The models that we introduce below are presented in a slightly different manner, but we show in Section 8 that they are equivalent. From the perspective of modal logic, an EHT model can be viewed as a refinement of S5 models (which are sets of valuations, alias sets of propositional variables) where valuations are replaced by HT models.

#### 4.1. The language of EHT

The language of EHT ( $\mathcal{L}_{\text{EHT}}$ ) is given by the following grammar:

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

where  $p$  ranges over the set  $\mathbb{P}$  of propositional variables.<sup>4</sup>

A finite set of EHT formulas is called an EHT *theory*, denoted by  $\Phi, \Psi, \dots$ . The set of propositional variables occurring in a formula  $\varphi$  is denoted by  $\mathbb{P}_\varphi$ . For example,  $\mathbb{P}_{K(p \rightarrow q)} = \{p, q\}$ . This generalises to EHT theories:  $\mathbb{P}_\Phi = \bigcup_{\varphi \in \Phi} \mathbb{P}_\varphi$ . As usual in HT,  $\top$ ,  $\neg\varphi$  and  $\varphi \leftrightarrow \psi$  respectively abbreviate  $\perp \rightarrow \perp$ ,  $\varphi \rightarrow \perp$  and  $(\varphi \rightarrow \psi) \wedge (\psi \rightarrow \varphi)$ . A formula is said to be *nonepistemic* (or *nonmodal*) if it contains neither  $K$  nor  $\hat{K}$ .

#### 4.2. EHT models

An EHT model has two components: the first is a non-empty collection of *there-valuations*, and the second assigns a *here-valuation* to each there-valuation in the first component. Formally, an EHT model is an ordered pair  $(\mathcal{T}, h)$  where

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

We call  $h$  *here-function* since it associates a here-valuation to each there-valuation in its domain. An EHT model  $(\mathcal{T}, h)$  can alternatively be described as a collection  $\{(h(T), T)\}_{T \in \mathcal{T}}$  of HT models: the *here-function*  $h$  implicitly determines HT models  $(h(T), T)$  for every  $T \in \mathcal{T}$ . In particular, when  $\mathcal{T} = \{T\}$  is a singleton, the EHT model  $(\mathcal{T}, h)$  can be identified with the HT model  $(h(T), T)$ . The inclusion constraint  $h(T) \subseteq T$ , for every  $T \in \mathcal{T}$  generalises the heredity constraint of HT to EHT.

We say that  $(\mathcal{T}, h)$  is *total on*  $\mathcal{S} \subseteq \mathcal{T}$  if  $h(T) = T$  for every  $T \in \mathcal{S}$ . If  $(\mathcal{T}, h)$  is total on  $\mathcal{T}$  then  $h$  is the identity function *id*, that is,  $h(T) = T$  for every  $T \in \mathcal{T}$ . We identify  $(\mathcal{T}, id)$  with the classical S5 model  $\mathcal{T} \subseteq 2^{\mathbb{P}}$ .

A *multi-pointed EHT model* is a pair  $((\mathcal{T}, h), \mathcal{T}_0)$  in which  $(\mathcal{T}, h)$  is an EHT model and  $\mathcal{T}_0 \subseteq \mathcal{T}$  is the non-empty set of *designated* (actual) worlds. When  $\mathcal{T}_0 = \{T_0\}$  we call it a *single-pointed EHT model*. We display multi-pointed models as collections  $\{(h(T), T)\}_{T \in \mathcal{T}}$  where the designated worlds are underlined. For example, we write  $\{(\emptyset, \{p\}), \underline{(\emptyset, \{q\})}\}$  for the single-pointed EHT model  $((\mathcal{T}, h), \{T_0\})$  where  $\mathcal{T} = \{\{p\}, \{q\}\}$ ,  $T_0 = \{q\}$  and  $h(\{p\}) = h(\{q\}) = \emptyset$ .

**Remark 1.** As  $h$  is a function, not every set of HT models corresponds to an EHT model. For example,  $\{(\emptyset, \{p\}), (\{p\}, \{p\})\}$  has no EHT counterpart. See Section 8 for more on this.

#### 4.3. EHT truth conditions

We now define the truth conditions for EHT formulas. Those for  $\perp$ ,  $\wedge$  and  $\vee$  are standard.

$$\begin{array}{ll} (\mathcal{T}, h), T \models_{\text{EHT}} p & \text{if } p \in h(T); \\ (\mathcal{T}, h), T \models_{\text{EHT}} \varphi \rightarrow \psi & \text{if } (\mathcal{T}, h), T \models_{\text{EHT}} \varphi \text{ implies } (\mathcal{T}, h), T \models_{\text{EHT}} \psi \text{ and} \\ & (\mathcal{T}, id), T \models_{\text{EHT}} \varphi \text{ implies } (\mathcal{T}, id), T \models_{\text{EHT}} \psi; \\ (\mathcal{T}, h), T \models_{\text{EHT}} K\varphi & \text{if } (\mathcal{T}, h), T' \models_{\text{EHT}} \varphi \text{ for every } T' \in \mathcal{T}; \\ (\mathcal{T}, h), T \models_{\text{EHT}} \hat{K}\varphi & \text{if } (\mathcal{T}, h), T' \models_{\text{EHT}} \varphi \text{ for some } T' \in \mathcal{T}. \end{array}$$

It follows from the definition of  $\neg\varphi$  as  $\varphi \rightarrow \perp$  that  $(\mathcal{T}, h), T \models_{\text{EHT}} \neg\varphi$  if and only if  $(\mathcal{T}, h), T \not\models_{\text{EHT}} \varphi$  and  $(\mathcal{T}, id), T \not\models_{\text{EHT}} \varphi$ . This truth condition will be further simplified in Item 1 of Proposition 3 below.

<sup>4</sup> As said in the introduction and as we will formally establish in Section 4.4, we prefer  $\hat{K}$  rather than Gelfond’s  $M$  operator because  $\hat{K}$  and  $M$  are interpreted in a different way: Gelfond’s  $M\varphi$  is equivalent to  $\text{not } K \text{ not } \varphi$ , while our  $\hat{K}\varphi$  is not equivalent to  $\neg K \neg\varphi$ .



For example,  $(\mathcal{T}, \mathfrak{h}), T_0 \models_{\text{EHT}} p \vee \neg p$  if and only if  $p \in \mathfrak{h}(T_0)$  or  $p \notin T_0$ . Moreover,  $(\mathcal{T}, \mathfrak{h}), T_0 \models_{\text{EHT}} \mathsf{K}(p \vee \neg p) \rightarrow \hat{\mathsf{K}}\neg p$  if and only if  $p \in T$  for some  $T \in \mathcal{T}$ . Finally,  $(\mathcal{T}, \mathfrak{h}), T \models_{\text{EHT}} \neg \mathsf{K}\neg \varphi \rightarrow \neg \mathsf{K}\varphi$  for every EHT model  $(\mathcal{T}, \mathfrak{h})$  and every  $T \in \mathcal{T}$ .

Given an EHT model  $(\mathcal{T}, \mathfrak{h})$  and a set  $\mathcal{T}_0 \subseteq \mathcal{T}$  of designated worlds, if we have  $(\mathcal{T}, \mathfrak{h}), T \models_{\text{EHT}} \varphi$  for every  $T \in \mathcal{T}_0$ , then we write  $(\mathcal{T}, \mathfrak{h}), \mathcal{T}_0 \models_{\text{EHT}} \varphi$  for short. We call  $((\mathcal{T}, \mathfrak{h}), \mathcal{T}_0)$  a multipointed EHT model of  $\varphi$ . Finally, we write  $(\mathcal{T}, \mathfrak{h}), \mathcal{T}_0 \models_{\text{EHT}} \Phi$  when  $(\mathcal{T}, \mathfrak{h}), \mathcal{T}_0 \models_{\text{EHT}} \varphi$  for every  $\varphi \in \Phi$ . Here are some examples:

1.  $\{(\emptyset, \{p\}), (\emptyset, \{q\})\} \models_{\text{EHT}} \neg p$  as  $\{(\emptyset, \{p\}), (\emptyset, \{q\})\} \not\models_{\text{EHT}} p$  and  $\{(\{p\}, \{p\}), (\{q\}, \{q\})\} \not\models_{\text{EHT}} p$ .
2.  $\{(\emptyset, \{p\}), (\emptyset, \{q\})\} \models_{\text{EHT}} \neg \neg p$  as  $\{(\{p\}, \{p\}), (\{q\}, \{q\})\} \models_{\text{EHT}} p$ .
3.  $\{(\emptyset, \{p\}), (\emptyset, \{q\})\} \models_{\text{EHT}} \mathsf{K}\neg r$  since  $\{(\emptyset, \{p\}), (\emptyset, \{q\})\} \models_{\text{EHT}} \neg r$ .
4.  $\{(\emptyset, \{p\}), (\emptyset, \{q\})\} \models_{\text{EHT}} \neg \neg \mathsf{K}(p \vee q)$  since  $\{(\{p\}, \{p\}), (\{q\}, \{q\})\} \models_{\text{EHT}} p \vee q$ .
5.  $\{(\emptyset, \{p\}), (\emptyset, \{q\})\} \models_{\text{EHT}} \hat{\mathsf{K}}\neg p$  since  $\{(\emptyset, \{p\}), (\emptyset, \{q\})\} \models_{\text{EHT}} \neg p$ .
6.  $\{(\emptyset, \{p\}), (\emptyset, \{q\})\} \models_{\text{EHT}} \neg \mathsf{K}\neg p$  since  $\{(\{p\}, \{p\}), (\{q\}, \{q\})\} \not\models_{\text{EHT}} \neg p$ .
7.  $\{(\emptyset, \emptyset)\} \models_{\text{EHT}} \neg \mathsf{K}\neg p \rightarrow p$  since  $\{(\emptyset, \emptyset)\} \not\models_{\text{EHT}} \neg \mathsf{K}\neg p$ .
8.  $\{(\{p\}, \{p\})\} \models_{\text{EHT}} \neg \mathsf{K}\neg p \rightarrow p$ .
9.  $\{(\emptyset, \{p\})\} \not\models_{\text{EHT}} \neg \mathsf{K}\neg p \rightarrow p$  since  $\{(\emptyset, \{p\})\} \models_{\text{EHT}} \neg \mathsf{K}\neg p$ , but  $\{(\emptyset, \{p\})\} \not\models_{\text{EHT}} p$ .

Observe that the satisfaction of the formulas of the form  $\mathsf{K}\varphi$  and  $\hat{\mathsf{K}}\varphi$  does not depend on designated worlds:  $(\mathcal{T}, \mathfrak{h}), T \models_{\text{EHT}} \mathsf{K}\varphi$  for some  $T \in \mathcal{T}$  if and only if  $(\mathcal{T}, \mathfrak{h}), T \models_{\text{EHT}} \mathsf{K}\varphi$  for every  $T \in \mathcal{T}$ . Also note that the satisfaction of the formulas of the form  $\neg \varphi$  is independent of the here-function  $h$ . Moreover, the satisfaction of the formulas of the form  $\neg \mathsf{K}\varphi$ ,  $\neg \neg \mathsf{K}\varphi$ ,  $\neg \hat{\mathsf{K}}\varphi$  and  $\neg \neg \hat{\mathsf{K}}\varphi$  depends on neither  $\mathfrak{h}$  nor designated worlds. We state this formally in Proposition 1 and Proposition 3 below.

**Proposition 1.** *The following are equivalent.*

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

**Proof.** This follows immediately from the fact that the satisfaction of a formula  $\mathsf{K}\varphi$  is independent of the designated worlds.  $\square$

The following result is the so-called heredity (monotonicity) property of intermediate logics, and therefore in particular of EHT: if a formula has an EHT model, then it also has a total EHT model; in other words, it has a classical S5 model.

**Proposition 2.** *If  $(\mathcal{T}, \mathfrak{h}), T \models_{\text{EHT}} \varphi$  then  $(\mathcal{T}, \text{id}), T \models_{\text{EHT}} \varphi$  (i.e.,  $\mathcal{T}, T \models_{\text{S5}} \varphi$ ).*

The proof is by induction on  $\varphi$ .

We now list some useful properties; in particular we simplify the satisfaction of negated EHT formulas.

**Proposition 3.** *For an EHT model  $(\mathcal{T}, \mathfrak{h})$  and an EHT formula  $\varphi$ , we have:*

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

**Proof.**

1. Suppose that  $(\mathcal{T}, \mathfrak{h}), T \models_{\text{EHT}} \neg \varphi$ . Then, by the EHT truth conditions, we get  $(\mathcal{T}, \mathfrak{h}), T \not\models_{\text{EHT}} \varphi$  and  $(\mathcal{T}, \text{id}), T \not\models_{\text{EHT}} \varphi$ , which further imply  $(\mathcal{T}, \text{id}), T \not\models_{\text{EHT}} \varphi$ . Conversely, suppose that  $(\mathcal{T}, \text{id}), T \not\models_{\text{EHT}} \varphi$ . Then, from Proposition 2, we obtain that  $(\mathcal{T}, \mathfrak{h}), T \not\models_{\text{EHT}} \varphi$  and  $(\mathcal{T}, \text{id}), T \not\models_{\text{EHT}} \varphi$ . Again by the EHT truth conditions, we have  $(\mathcal{T}, \mathfrak{h}), T \models_{\text{EHT}} \neg \varphi$ .

2. Put  $\neg \varphi$  in place of  $\varphi$  in Proposition 3.1, and the result follows:  $(\mathcal{T}, \mathfrak{h}), T \models_{\text{EHT}} \neg \neg \varphi$  if and only if  $(\mathcal{T}, \text{id}), T \not\models_{\text{EHT}} \neg \varphi$  if and only if  $(\mathcal{T}, \text{id}), T \models_{\text{EHT}} \varphi$ .

3. Put  $\mathsf{K}\varphi$  in place of  $\varphi$  in Proposition 3.1, and the result follows:  $(\mathcal{T}, \mathfrak{h}), T \models_{\text{EHT}} \neg \mathsf{K}\varphi$  if and only if  $(\mathcal{T}, \text{id}), T \not\models_{\text{EHT}} \mathsf{K}\varphi$  if and only if  $(\mathcal{T}, \text{id}), T' \not\models_{\text{EHT}} \varphi$  for some  $T' \in \mathcal{T}$ .

**Table 3**

Axiom schemas of intuitionistic S5 [30] (notation adapted).

0.	All substitution instances of theorems of intuitionistic propositional logic
1.	$K(\varphi \rightarrow \psi) \rightarrow (K\varphi \rightarrow K\psi)$
2.	$K(\varphi \rightarrow \psi) \rightarrow (\hat{K}\varphi \rightarrow \hat{K}\psi)$
3.	$\neg\hat{K}\perp$
4.	$\hat{K}(\varphi \vee \psi) \rightarrow (\hat{K}\varphi \vee \hat{K}\psi)$
5.	$(\hat{K}\varphi \rightarrow K\psi) \rightarrow K(\varphi \rightarrow \psi)$
6.	$\hat{K}\top$
7.	$(K\varphi \rightarrow \varphi) \wedge (\varphi \rightarrow \hat{K}\varphi)$
8.	$(\hat{K}K\varphi \rightarrow \varphi) \wedge (\varphi \rightarrow K\hat{K}\varphi)$
9.	$(K\varphi \rightarrow KK\varphi) \wedge (\hat{K}\hat{K}\varphi \rightarrow \hat{K}\varphi)$
10.	$(\hat{K}K\varphi \rightarrow K\varphi) \wedge (\hat{K}\varphi \rightarrow K\hat{K}\varphi)$

4. Put  $\hat{K}\varphi$  in place of  $\varphi$  in Proposition 3.1, and the result follows:  $(\mathcal{T}, \mathfrak{h}), T \models_{\text{EHT}} \neg\hat{K}\varphi$  if and only if  $(\mathcal{T}, id), T \not\models_{\text{EHT}} \hat{K}\varphi$  if and only if  $(\mathcal{T}, id), T' \not\models_{\text{EHT}} \varphi$  for every  $T' \in \mathcal{T}$ .  $\square$

We end with an appropriate definition of bisimilarity. Let  $(\mathcal{T}_1, \mathfrak{h}_1)$  and  $(\mathcal{T}_2, \mathfrak{h}_2)$  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,<sup>5</sup> and
2. if  $T_1 Z T_2$  then  $T_1 \cap P = T_2 \cap P$  and  $h_1(T_1) \cap P = h_2(T_2) \cap P$ .

If there exists a  $P$ -bisimulation  $Z$  between  $(\mathcal{T}_1, h_1)$  and  $(\mathcal{T}_2, h_2)$  such that  $T_1 Z T_2$  then the pointed models  $((\mathcal{T}_1, h_1), T_1)$  and  $((\mathcal{T}_2, h_2), T_2)$  are called  $P$ -bisimilar.

**Proposition 4.** Let  $\varphi$  be an EHT formula. Let  $((\mathcal{T}_1, \mathfrak{h}_1), T_1)$  and  $((\mathcal{T}_2, h_2), T_2)$  be two single-pointed EHT models such that  $((\mathcal{T}_1, h_1), T_1)$  and  $((\mathcal{T}_2, h_2), T_2)$  are  $\mathbb{P}_\varphi$ -bisimilar. Then, we have

$$(\mathcal{T}_1, \mathfrak{h}_1), T_1 \models_{\text{EHT}} \varphi \text{ if and only if } (\mathcal{T}_2, \mathfrak{h}_2), T_2 \models_{\text{EHT}} \varphi.$$

**Proof.** Let  $((\mathcal{T}_1, h_1), T_1)$  and  $((\mathcal{T}_2, h_2), T_2)$  be two single-pointed EHT models such that  $T_1$  and  $T_2$  are  $\mathbb{P}_\varphi$ -bisimilar. Then by definition there exists a  $\mathbb{P}_\varphi$ -bisimulation  $Z \subseteq \mathcal{T}_1 \times \mathcal{T}_2$  satisfying:  $T_1 \cap \mathbb{P}_\varphi = T_2 \cap \mathbb{P}_\varphi$  and  $h_1(T_1) \cap \mathbb{P}_\varphi = h_2(T_2) \cap \mathbb{P}_\varphi$ . So, the HT models  $(\mathfrak{h}_1(T_1), T_1)$  and  $(\mathfrak{h}_2(T_2), T_2)$  agree on the propositional variables of  $\varphi$ . Moreover,  $Z$  and  $Z^{-1}$  are serial. So, for every  $T \in \mathcal{T}_1$ , there is  $T' \in \mathcal{T}_2$  such that the here and there-valuations of  $(h_1(T), T)$  and  $(h_2(T'), T')$  agree on  $\mathbb{P}_\varphi$ . The same reasoning also applies the other way round. Then the result follows by induction on  $\varphi$ .  $\square$

#### 4.4. EHT validity

A formula  $\varphi$  is called EHT satisfiable if  $(\mathcal{T}, \mathfrak{h}), T \models_{\text{EHT}} \varphi$  for some EHT model  $(\mathcal{T}, \mathfrak{h})$  and  $T \in \mathcal{T}$ . Then,  $\varphi$  is EHT valid if  $(\mathcal{T}, \mathfrak{h}), \mathcal{T} \models_{\text{EHT}} \varphi$  for every EHT model  $(\mathcal{T}, \mathfrak{h})$ .

All principles of intuitionistic modal logics that were studied in the literature [31,28,30] are EHT valid. In particular, the axiom schemas for intuitionistic S5 [30] are all valid; they are listed in Table 3. The inference rules are modus ponens and necessitation. As always in intuitionistic modal logics,  $K$  and  $\hat{K}$  are not dual: while  $\hat{K}\varphi \rightarrow \neg K\neg\varphi$  is valid, the other direction is not. Here are some more examples:  $K\varphi \rightarrow \neg\hat{K}\neg\varphi$ ,  $\neg\hat{K}\neg\varphi \rightarrow \neg\neg\hat{K}\varphi$  and  $\neg\neg K\varphi \rightarrow \neg K\neg\varphi$  are all valid while their converses are not. On the other hand, none of  $\neg\neg K\varphi \rightarrow K\varphi$  and  $K\neg\neg\varphi \rightarrow K\varphi$  is EHT valid. The same holds if we replace  $K$  by  $\hat{K}$ .<sup>6</sup> However:

**Proposition 5.** The equivalences  $\neg K\varphi \leftrightarrow \hat{K}\neg\varphi$  and  $\neg\hat{K}\varphi \leftrightarrow K\neg\varphi$  are EHT valid.

**Proof.** Given an EHT model  $(\mathcal{T}, \mathfrak{h})$  and  $\mathcal{T}_0 \subseteq \mathcal{T}$ , we have:

$$\begin{aligned} (\mathcal{T}, \mathfrak{h}), \mathcal{T}_0 \models_{\text{EHT}} \hat{K}\neg\varphi & \text{ iff } (\mathcal{T}, \mathfrak{h}), T_0 \models_{\text{EHT}} \neg\varphi \text{ for some } T_0 \in \mathcal{T} \\ & \text{ iff } (\mathcal{T}, id), T_0 \not\models_{\text{EHT}} \varphi \text{ for some } T_0 \in \mathcal{T} \text{ (by Proposition 3.1)} \\ & \text{ iff } (\mathcal{T}, \mathfrak{h}), \mathcal{T}_0 \models_{\text{EHT}} \neg K\varphi \text{ (by Proposition 3.3).} \end{aligned}$$

<sup>5</sup> The inverse relation  $Z^{-1} \subseteq \mathcal{T}_2 \times \mathcal{T}_1$  of a binary relation  $Z \subseteq \mathcal{T}_1 \times \mathcal{T}_2$  is  $Z^{-1} = \{(y, x) : (x, y) \in Z\}$ . The relation  $Z \subseteq \mathcal{T}_1 \times \mathcal{T}_2$  is serial if for every  $T_1 \in \mathcal{T}_1$  there is  $T_2 \in \mathcal{T}_2$  such that  $(T_1, T_2) \in Z$ .

<sup>6</sup> Take  $\varphi = p$ . Then the EHT model  $\{(\emptyset, \{p\})\}$  provides a countermodel for all examples above except  $\neg\neg\hat{K}p \rightarrow \neg\hat{K}\neg p$  and  $\neg K\neg p \rightarrow \neg\neg Kp$ . For these two,  $\{(\emptyset, \emptyset), (\emptyset, \{p\})\}$  works as a counterexample.

$$\begin{aligned}
(\mathcal{T}, \mathfrak{h}), \mathcal{T}_0 \models_{\text{EHT}} \mathsf{K}\neg\varphi & \text{ iff } (\mathcal{T}, \mathfrak{h}), T \models_{\text{EHT}} \neg\varphi \text{ for every } T \in \mathcal{T} \\
& \text{ iff } (\mathcal{T}, id), T \not\models_{\text{EHT}} \varphi \text{ for every } T \in \mathcal{T} \text{ (by Proposition 3.1)} \\
& \text{ iff } (\mathcal{T}, \mathfrak{h}), \mathcal{T}_0 \models_{\text{EHT}} \neg\hat{\mathsf{K}}\varphi \text{ (by Proposition 3.4).}
\end{aligned}$$

Consequently,  $\hat{\mathsf{K}}\neg\varphi$  and  $\neg\mathsf{K}\varphi$  are logically equivalent; so are  $\mathsf{K}\neg\varphi$  and  $\neg\hat{\mathsf{K}}\varphi$ .  $\square$

As an immediate corollary of this proposition, we also have:

**Corollary 1.** *The equivalences  $\neg\neg\mathsf{K}\varphi \leftrightarrow \mathsf{K}\neg\neg\varphi$  and  $\neg\neg\hat{\mathsf{K}}\varphi \leftrightarrow \hat{\mathsf{K}}\neg\neg\varphi$  are EHT valid.*

**Proof.** By Proposition 5,  $\neg\mathsf{K}\varphi \leftrightarrow \hat{\mathsf{K}}\neg\varphi$  is EHT valid; then, so is  $\neg\neg\mathsf{K}\varphi \leftrightarrow \neg\hat{\mathsf{K}}\neg\varphi$  (\*). Again by Proposition 5,  $\neg\hat{\mathsf{K}}\varphi \leftrightarrow \mathsf{K}\neg\varphi$  is EHT valid; put  $\neg\varphi$  in place of  $\varphi$  and then we get:  $\neg\hat{\mathsf{K}}\neg\varphi \leftrightarrow \mathsf{K}\neg\neg\varphi$  (\*\*). Thus, from (\*) and (\*\*) we obtain that  $\neg\neg\mathsf{K}\varphi \leftrightarrow \mathsf{K}\neg\neg\varphi$  is EHT valid. Note that  $\mathsf{K}$  and  $\hat{\mathsf{K}}$  are symmetric in Proposition 5, so the second result follows similarly to the first result, and we also have:  $\neg\neg\hat{\mathsf{K}}\varphi \leftrightarrow \hat{\mathsf{K}}\neg\neg\varphi$  is EHT valid.  $\square$

## 5. Epistemic and autoepistemic equilibrium models

Pearce defined equilibrium models (EMs) of an HT formula as its classical models satisfying a minimality condition when viewed as total HT models. Formally,  $T \subseteq \mathbb{P}$  is an equilibrium model of an HT formula  $\varphi$  if  $T$  classically satisfies  $\varphi$  (i.e.,  $T \models \varphi$ ) and any HT model  $(H, T)$  does not satisfy  $\varphi$  (i.e.,  $H, T \not\models_{\text{HT}} \varphi$ ) when  $H \neq T$ . In a sense, truth is minimised: to witness, the only equilibrium model of  $p \rightarrow p$  is  $\emptyset$ . Our epistemic equilibrium models (EEMs) generalise Pearce's equilibrium models from classical models to S5 models in a natural way. The resulting nonmonotonic consequence relation is a conservative extension of the standard equilibrium consequence relation: they coincide for the fragment of non-epistemic formulas. However, such models (EEMs) only minimise truth, while the semantics of epistemic specifications involves also a minimisation of knowledge. In the second part of the section, we therefore define autoepistemic equilibrium models (AEEMs) as EEMs that are maximal under some orderings that maximise ignorance.

### 5.1. Total models and their weakening

Remember that an EHT model  $(\mathcal{T}, \mathfrak{h})$  is total when  $h = id$ , where  $id$  refers to the identity function (see Section 4.2). A total EHT model corresponds to a classical S5 model; so validity in classical S5 models is the same as validity in total EHT models. We can therefore identify  $(\mathcal{T}, id), \mathcal{T}_0 \models_{\text{EHT}} \varphi$  with  $\mathcal{T}, \mathcal{T}_0 \models_{\text{S5}} \varphi$ . Given two EHT models  $(\mathcal{T}_1, h_1)$  and  $(\mathcal{T}_2, h_2)$ , we say that  $(\mathcal{T}_1, h_1)$  is *weaker* than  $(\mathcal{T}_2, h_2)$  if the there-sets  $\mathcal{T}_1$  and  $\mathcal{T}_2$  are identical and all the here-sets  $h_1(T)$  are pointwise included in the here-sets  $h_2(T)$ . Formally we write

$$(\mathcal{T}_1, \mathfrak{h}_1) \leq (\mathcal{T}_2, \mathfrak{h}_2)$$

if  $\mathcal{T}_1 = \mathcal{T}_2$  and  $h_1(T) \subseteq h_2(T)$ , for every  $T \in \mathcal{T}_1$ .

This is a non-strict partial order. The corresponding strict partial order is defined in the standard way:

$$(\mathcal{T}_1, \mathfrak{h}_1) \triangleleft (\mathcal{T}_2, \mathfrak{h}_2) \text{ if } (\mathcal{T}_1, \mathfrak{h}_1) \leq (\mathcal{T}_2, \mathfrak{h}_2) \text{ and } (\mathcal{T}_2, \mathfrak{h}_2) \not\leq (\mathcal{T}_1, \mathfrak{h}_1)$$

and we say that  $(\mathcal{T}_1, \mathfrak{h}_1)$  is *strictly weaker* than  $(\mathcal{T}_2, \mathfrak{h}_2)$ .

### 5.2. Epistemic equilibrium models

We have observed that a total EHT model  $(\mathcal{T}, id)$  can be identified with the classical S5 model  $\mathcal{T}$ . We now define the *epistemic equilibrium models* (EEMs) of  $\varphi$  as particular S5 models:

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

In other words, an EEM of  $\varphi$  is a minimal total EHT model of  $\varphi$  with respect to the ordering  $\leq$ . So, the *minimality condition* requires that there is no EHT model  $(\mathcal{T}, \mathfrak{h})$  of  $\varphi$  that is strictly weaker than  $(\mathcal{T}, id)$ . It follows that all EHT valid formulas have exactly one EEM, namely  $\{\emptyset\}$ . For example, the unique EEM of the atomic  $p$  is  $\{\{p\}\}$ . This is also the unique EEM of  $\mathsf{K}p$ . The EEMs of  $\hat{\mathsf{K}}p$  are  $\{\{p\}\}$  and  $\{\emptyset, \{p\}\}$ . The formulas  $\neg\mathsf{K}\neg p$  and  $\neg\hat{\mathsf{K}}\neg p$  have no EEM.

**Proposition 6.** *Let  $\varphi$  be an EHT formula.*

1. *For every  $\mathcal{T} \in EEM(\varphi)$  we have  $\mathcal{T} \subseteq 2^{\mathbb{P}_\varphi}$ .*
2. *If  $\{\emptyset\} \models_{\text{S5}} \varphi$  then  $EEM(\neg\neg\varphi) = \{\{\emptyset\}\}$  and  $EEM(\neg\varphi) = \emptyset$ . Otherwise, if  $\{\emptyset\} \not\models_{\text{S5}} \varphi$ , then  $EEM(\neg\neg\varphi) = \emptyset$  and  $EEM(\neg\varphi) = \{\{\emptyset\}\}$ .*

3.  $EEM(K\varphi) = EEM(\varphi)$ .

**Proof.** Let  $\varphi$  be an EHT formula.

2. Assume that  $\{\emptyset\} \models_{SS} \varphi$ . Then  $\{\emptyset\} \models_{SS} \neg\neg\varphi$  since double negation vanishes in classical S5. The minimality condition for EEMs is trivially satisfied for  $\mathcal{T} = \{\emptyset\}$ . As a result,  $\{\emptyset\} \in EEM(\neg\neg\varphi)$ . By Proposition 3.2, we know that  $(\mathcal{T}, h), T \models_{EHT} \neg\neg\varphi$  if and only if  $(\mathcal{T}, id), T \models_{EHT} \varphi$  if and only if  $(\mathcal{T}, id), T \models_{EHT} \neg\neg\varphi$ . So, for every  $\mathcal{T} \neq \{\emptyset\}$  and every  $h \neq id$  such that  $(\mathcal{T}, h) \triangleleft (\mathcal{T}, id)$ , when  $\mathcal{T}, T \models_{SS} \neg\neg\varphi$ , then  $(\mathcal{T}, h), T \models_{EHT} \neg\neg\varphi$ . So, the minimality condition of EEMs always fails for  $\mathcal{T} \neq \{\emptyset\}$  and the EHT formula  $\neg\neg\varphi$ . Consequently,  $\{\emptyset\}$  is the unique element of  $EEM(\neg\neg\varphi)$ . Moreover, by Proposition 3.1, we also know that  $(\mathcal{T}, h), T \models_{EHT} \neg\varphi$  if and only if  $(\mathcal{T}, id), T \not\models_{EHT} \varphi$  if and only if  $(\mathcal{T}, id), T \models_{EHT} \neg\varphi$ . Thus, again for every  $\mathcal{T} \neq \{\emptyset\}$  and every  $h \neq id$  such that  $(\mathcal{T}, h) \triangleleft (\mathcal{T}, id)$ , when  $\mathcal{T}, T \models_{SS} \neg\varphi$ ,  $(\mathcal{T}, h), T \models_{EHT} \neg\varphi$ . So, the minimality condition again always fails for  $\mathcal{T} \neq \{\emptyset\}$  and the EHT formula  $\neg\varphi$ . However, by assumption, we have  $\{\emptyset\} \models_{SS} \varphi$ , that is,  $\{\emptyset\} \not\models_{SS} \neg\varphi$ . Thus, the only candidate  $\{\emptyset\}$  is not an EEM of  $\neg\varphi$ . As a result,  $EEM(\neg\varphi) = \emptyset$ . On the other hand, when  $\{\emptyset\} \not\models_{SS} \varphi$ , that is,  $\{\emptyset\} \models_{SS} \neg\varphi$ , we simply replace  $\varphi$  by  $\neg\varphi$  in the results above. So we have  $EEM(\neg\varphi) = \{\{\emptyset\}\}$  (recall that  $\neg\neg\neg\varphi \leftrightarrow \neg\varphi$  is EHT valid) and  $EEM(\neg\neg\varphi) = \emptyset$ .

3. Let  $\mathcal{T} \in EEM(K\varphi)$  be arbitrary. Then  $\mathcal{T}, T \models_{SS} K\varphi$ , that is,  $\mathcal{T}, T \models_{SS} \varphi$ . Moreover, the minimality condition implies that  $(\mathcal{T}, h), T \not\models_{EHT} K\varphi$  for every  $h \neq id$ . So, for any  $h \neq id$ ,  $(\mathcal{T}, h), T' \not\models_{EHT} \varphi$  for some  $T' \in \mathcal{T}$ , i.e.,  $(\mathcal{T}, h), T \not\models_{EHT} \varphi$ . As a result,  $\mathcal{T} \in EEM(\varphi)$ . The opposite direction works in a similar way. Hence,  $EEM(\varphi) = EEM(K\varphi)$ .  $\square$

**Proposition 7.** Let  $\varphi$  be a nonepistemic EHT formula (i.e., an HT formula). Let  $EM(\varphi)$  be the set of all (classical) equilibrium models of  $\varphi$ . Then, we have:

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

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

**Proof.** Let  $\varphi \in \mathcal{L}_{EHT}$  be non-epistemic. Let  $EM(\varphi)$  be the set of all equilibrium models of  $\varphi$ . Let  $\mathcal{T} \subseteq EM(\varphi)$  be non-empty. Since every  $T \in \mathcal{T}$  is an equilibrium model of  $\varphi$ , for every  $T \in \mathcal{T}$  we have (i)  $T \models \varphi$  and (ii)  $h(T), T \not\models_{HT} \varphi$  for every  $h \neq id$  such that  $h(T) \subset T$ . Thus, from (i) we obtain that  $\mathcal{T}, T \models_{SS} \varphi$ , and by (ii) we get  $(\mathcal{T}, h), T \not\models_{EHT} \varphi$ , for every  $h \neq id$  such that  $(\mathcal{T}, h) \triangleleft (\mathcal{T}, id)$ . As a result,  $\mathcal{T} \in EEM(\varphi)$ . Conversely, let  $\mathcal{T}' \in EEM(\varphi)$  then by definition, (i)  $\mathcal{T}', T' \models_{SS} \varphi$  and (ii) for every  $h \neq id$ ,  $(\mathcal{T}', h), T' \not\models_{EHT} \varphi$ . From (i) we obtain that for every  $T' \in \mathcal{T}'$ ,  $T' \models \varphi$  and from (ii) we obtain that  $h(T'), T' \not\models_{HT} \varphi$  for every  $h \neq id$ . As a result, for every  $T' \in \mathcal{T}'$ ,  $T' \in EM(\varphi)$ , and so  $\mathcal{T}' \subseteq EM(\varphi)$ . Thus, we conclude that  $EEM(\varphi) = \{\mathcal{T} \subseteq EM(\varphi) : \mathcal{T} \neq \emptyset\}$ .

Let  $\mathcal{T} \in EEM(\hat{K}\varphi)$  be arbitrary. Then by definition, we have: (i)  $\mathcal{T}, T \models_{SS} \hat{K}\varphi$  and (ii)  $(\mathcal{T}, h), T \not\models_{EHT} \hat{K}\varphi$  for every  $h \neq id$ . From (i), we obtain that  $T_0 \models \varphi$  for some  $T_0 \in \mathcal{T}$  and from (ii), we obtain that  $(\mathcal{T}, h), T \not\models_{EHT} \varphi$  for every  $T \in \mathcal{T}$  and every  $h \neq id$ . Thus,  $(\mathcal{T}, h'), T_0 \not\models_{EHT} \varphi$  for every  $h'$  such that  $h'(T_0) \subset T_0$  and  $h'(T) = T$  for every  $T \in \mathcal{T} \setminus \{T_0\}$ . Since  $\varphi$  is non-epistemic, we have  $h'(T_0), T_0 \not\models_{HT} \varphi$  for such  $h'$ 's. As a result,  $T_0 \in EM(\varphi)$ . Once again recall that  $T_0 \in \mathcal{T}$ , then we have:

Case 1) let  $\emptyset \in EM(\varphi)$ , then  $\emptyset \models \varphi$ . Hence, we conclude that  $\mathcal{T} = \{T_0\}$ : otherwise, on the one hand, for any  $T' \in \mathcal{T}$  satisfying  $T' \neq \emptyset$  and  $T' \neq T_0$ , every here-function  $h$  such that  $h(T_0) = T_0$  falsifies the minimality condition (ii) above; on the other hand, for  $T' = \emptyset$  and  $T_0 \neq \emptyset$ , since  $\emptyset \models \varphi$ , any here-function  $h$  would falsify the minimality condition (ii) above.

Case 2) let  $\emptyset \notin EM(\varphi)$ , then  $\emptyset \not\models \varphi$  and  $T_0 \neq \emptyset$ . Following the same argument as above, we conclude that  $\mathcal{T} = \{T_0\}$  or  $\mathcal{T} = \{\emptyset, T_0\}$  (since  $\emptyset \not\models \varphi$ ).

The proof of the other direction of the set inclusion is similar.  $\square$

**Corollary 2.** For a nonepistemic EHT formula  $\varphi$ ,  $EEM(\varphi) = \emptyset$  iff  $EEM(\hat{K}\varphi) = \emptyset$ .

**Proof.** Let  $EEM(\varphi) = \emptyset$ . Then, assume for a contradiction that  $EEM(\hat{K}\varphi) \neq \emptyset$ . Thus, there is  $\mathcal{T} \in EEM(\hat{K}\varphi)$  satisfying: (i)  $\mathcal{T}, T \models_{SS} \hat{K}\varphi$  and (ii)  $(\mathcal{T}, h), T \not\models_{EHT} \hat{K}\varphi$  for every  $h \neq id$ . By (i), we have  $\mathcal{T}, T_0 \models_{SS} \varphi$  for some  $T_0 \in \mathcal{T}$ . By (ii), we have  $(\mathcal{T}, h), T \not\models_{EHT} \varphi$  for every  $T \in \mathcal{T}$  and for every  $h \neq id$ . By Proposition 7, the singleton  $\{T_0\}$  is an EEM of  $\varphi$ , which contradicts our initial assumption. Therefore,  $EEM(\hat{K}\varphi) = \emptyset$ . The proof of the opposite side is similar.  $\square$

We illustrate the above propositions by some examples. As  $EM(p \vee \neg p) = \{\emptyset, \{p\}\}$ , it follows from Proposition 7 that  $\hat{K}(p \vee \neg p)$  has the EEMs  $\{\emptyset\}$  and  $\{\{p\}\}$  and that  $p \vee \neg p$  has one more EEM, viz.  $\{\emptyset, \{p\}\}$ . Then, it follows from Item 3 of Proposition 6 that the EEMs of  $K(p \vee \neg p)$  are  $\{\emptyset\}$ ,  $\{\{p\}\}$  and  $\{\emptyset, \{p\}\}$ . Note that  $\neg\neg p$  has no EMs: the only candidate is  $\emptyset$ ,<sup>7</sup>

<sup>7</sup> In Proposition 3.2, if we consider  $\mathcal{T} = \{T\}$  and  $\varphi$  to be non-epistemic, then we have:  $H, T \models_{HT} \neg\neg\varphi$  if and only if  $T \models \varphi$  if and only if  $T \models \neg\neg\varphi$ , so for any  $T \neq \emptyset$ , the minimality condition does not hold.

**Table 4**

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

Formula	World views	EEMs
$Kp \rightarrow p$	$\{\emptyset\}$	$\{\emptyset\}$
$\neg K\neg p \rightarrow p$	$\{\{p\}\}$	$\{\emptyset\}$ and $\{\{p\}\}$
$\hat{K}p \rightarrow p$		$\{\emptyset\}$
$p \vee q$	$\{\{p\}, \{q\}\}$	$\{\{p\}\}$ and $\{\{q\}\}$ and $\{\{p\}, \{q\}\}$
$(p \vee q) \wedge (\neg K\neg q \rightarrow p)$	$\{\{p\}\}$	$\{\{p\}\}$
$(p \vee q) \wedge (Kp \rightarrow r)$	$\{\{p\}, \{q\}\}$	$\{\{p\}, \{q\}\}$ and $\{\{p, r\}\}$ and $\{\{q\}\}$

but  $\emptyset \not\models \neg\neg p$  (since  $\emptyset \not\models p$ ). So  $\neg\neg p$ ,  $\hat{K}\neg\neg p$  and  $K\neg\neg p$  have no EEMs (cf. Item 3 of Proposition 6 and Corollary 2) and neither do  $\neg\neg\hat{K}p$ ,  $\neg K\neg p$ ,  $\neg\neg Kp$  and  $\neg\hat{K}\neg p$  (cf. Proposition 5). Finally, while the EEMs of  $\varphi \in \mathcal{L}_{\text{HT}}$  and  $\hat{K}\varphi$  differ in general (cf. Item 3 of Proposition 6 and Proposition 7), those of  $\neg p$ ,  $\hat{K}\neg p$  and  $K\neg p$  coincide since it is only  $\{\emptyset\}$ . Moreover,  $\neg Kp$  and  $\neg K\neg p$  also have one EEM, viz.  $\{\emptyset\}$  (cf. Proposition 5).

Table 4 illustrates our definition by comparing the world views of the epistemic specifications of Table 2 with the EEMs of their EHT counterparts.<sup>8</sup> This table helps us realise that not every EEM of an epistemic specification is a world view. Moreover, we put an additional EHT formula  $\hat{K}p \rightarrow p$  to make the comparison of the modal operators  $\hat{K}$  and  $M$  (i.e.,  $\neg K\neg$ ) apparent. As  $\hat{K}p$  is not equivalent to  $\neg K\neg p$ , it has two EEMs  $\{\{p\}\}$  and  $\{\emptyset, \{p\}\}$  while  $\neg K\neg p$  has no EEMs. Since  $\hat{K}p$  is purely positive (not including negation), we find this result intuitive. The example  $\hat{K}p \rightarrow p$  will illustrate the benefit of having the  $\hat{K}$  operator as nondual of  $K$ : while  $\{\{p\}\}$  is an EEM of  $\neg K\neg p \rightarrow p$ , it is not an EEM for  $\hat{K}p \rightarrow p$  because  $(\emptyset, \{p\}) \models \hat{K}p \rightarrow p$  as well.

### 5.3. Autoepistemic equilibrium models

Although EEMs provide an interesting epistemic generalisation of EMs, they are somewhat too weak to provide an interesting semantics of epistemic specifications because they only minimise truth, but not knowledge. In order to select the intended models we take inspiration from autoepistemic logic [32] and the logic of all-that-I-know [33]: we are going to maximise ignorance and, by that, minimise knowledge. Consider the fourth formula  $\varphi_1 = p \vee q$  of Table 4. The EEM  $\{\{p\}, \{q\}\}$  contains each of the other EEMs and is therefore the inclusion-maximal EEM of  $p \vee q$ : it is the EEM where knowledge is minimal, and it matches the single world view of  $\{p \text{ or } q \leftarrow\}$ .

However, inclusion-maximality is sometimes not sufficient. Consider the second formula  $\neg K\neg p \rightarrow p$  of Table 4: while none of its EEMs is included in the other, one would like to eliminate  $\{\emptyset\}$  because it appears to contradict our intuitions about epistemic specifications.<sup>9</sup> Intuitively, the interpretation of  $K$  in EEMs should quantify over all possible answer sets. We consider the union of all EEMs of  $\varphi$ ,  $\bigcup EEM(\varphi)$ , to be the set of all candidate possible answer sets, and we select the preferred EEMs under set inclusion and a  $\varphi$ -indexed ordering over S5 models that is determined by  $\bigcup EEM(\varphi)$ .

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  $\mathcal{T}_0$ :

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

The first condition requires that for the total multipointed EHT model  $((\mathcal{T}, \text{id}), \mathcal{T}_0)$ , we have  $((\mathcal{T}, \text{id}), \mathcal{T}_0) \models_{\text{EHT}} \varphi$ . The second *minimality of truth* condition requires that there is no EHT model  $((\mathcal{T}, \hat{h}), \mathcal{T}_0)$  of  $\varphi$  strictly weaker than  $((\mathcal{T}, \text{id}), \mathcal{T}_0)$  on  $\mathcal{T}_0$ . By the latter we understand that  $\hat{h}(T) \subset T$  for some  $T \in \mathcal{T}_0$  and  $\hat{h}(T) = T$  for every  $T \in \mathcal{T} \setminus \mathcal{T}_0$ .

EEMs can be defined in terms of  $\models^*$ :

**Remark 2.** For every EHT formula  $\varphi$ ,  $EEM(\varphi) = \{\mathcal{T} \subseteq 2^{\mathbb{P}} : \mathcal{T}, \mathcal{T} \models^* \varphi\}$ .

We take up the above examples  $\neg K\neg p \rightarrow p$  and  $\hat{K}p \rightarrow p$  to illustrate our definition (remember that the designated worlds are underlined):

- $\{\{p\}, \underline{\emptyset}\} \not\models^* \neg K\neg p \rightarrow p$  because  $\{\{p\}, \underline{\emptyset}\} \not\models_{\text{S5}} \neg K\neg p \rightarrow p$ ;
- $\{\underline{\{p\}}, \emptyset\} \models^* \neg K\neg p \rightarrow p$  because  $\{\underline{\{p\}}, \emptyset\} \models_{\text{S5}} \neg K\neg p \rightarrow p$  and  $\{\underline{\emptyset}, \{p\}\}, (\emptyset, \emptyset) \not\models_{\text{EHT}} \neg K\neg p \rightarrow p$ ;

<sup>8</sup> The translation is straightforward; it is defined formally in Section 6.1. Also notice that  $\hat{K}$  has no counterpart in epistemic specifications.

<sup>9</sup> Kahl, in his PhD thesis, (see pages 12 and 25 of [6]) argues that at the very least, a rational agent should not accept both  $\{\emptyset\}$  and  $\{\{p\}\}$  together as the world views of  $p \leftarrow Mp$ . However, the question of which of them is intuitive is still under discussion although the majority finds  $\{\{p\}\}$  more intuitive. We choose to follow Kahl's reasoning: according to Kahl's preference relation on literals, it is easier to accept  $Mp$  (lower conviction) compared to  $p$ , so we expect to see  $\{\{p\}\}$  as a world view of  $p \leftarrow Mp$ .

- $\{\{p\}, \emptyset\} \not\models^* \hat{K}p \rightarrow p$  because  $(\emptyset, \{p\}), (\emptyset, \emptyset) \models_{\text{EHT}} \hat{K}p \rightarrow p$ .

The next step in our construction is to define a *formula-indexed preorder* over S5 models. The intuition behind this order is to check and compare the behaviour of EEMs with respect to each possible answer set (or belief set) candidate in  $\bigcup EEM(\varphi)$ . Let:

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

The strict version of  $\leq_{\varphi}$  is defined in the standard way:  $\mathcal{T} <_{\varphi} \mathcal{S}$  if and only if  $\mathcal{T} \leq_{\varphi} \mathcal{S}$  and  $\mathcal{S} \not\leq_{\varphi} \mathcal{T}$ . When  $\mathcal{T} <_{\varphi} \mathcal{S}$ , we say that  $\mathcal{S}$  is *preferred* over  $\mathcal{T}$  with respect to the ordering  $\leq_{\varphi}$ . When  $\mathcal{T} \leq_{\varphi} \mathcal{S}$  and  $\mathcal{S} \leq_{\varphi} \mathcal{T}$ , we say that  $\mathcal{S}$  is *equivalent* to  $\mathcal{T}$  with respect to  $\leq_{\varphi}$ ; we denote this by  $\mathcal{T} \approx_{\varphi} \mathcal{S}$ .

We are then interested in EEMs of  $\varphi$  that are maximal with respect to set inclusion  $\subseteq$  and the preference ordering  $\leq_{\varphi}$ . We say that  $\mathcal{T}$  is an *autoepistemic equilibrium model* (AEEM) of  $\varphi$  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  $\mathcal{T} <_{\varphi} \mathcal{S}$ .

The second condition corresponds to the maximal ignorance condition in the definition of world views of Section 3.2. The third condition says that when  $\mathcal{T} <_{\varphi} \mathcal{S}$  then  $\mathcal{S}$  can accommodate more relevant possibilities than  $\mathcal{T}$ , where a relevant possibility is a valuation occurring in some EEM of  $\varphi$ .

For example,  $\varphi_1 = p \vee q$  has three EEMs  $\{\{p\}\}$ ,  $\{\{q\}\}$  and  $\{\{p\}, \{q\}\}$ . While they are all equivalent with respect to  $\leq_{\varphi_1}$ , the last is the only AEEM of  $\varphi_1$  because it is maximal with respect to set inclusion. Similarly,  $\varphi_2 = \hat{K}(p \vee q)$  has four EEMs  $\{\{p\}\}$ ,  $\{\{q\}\}$ ,  $\{\emptyset, \{p\}\}$  and  $\{\emptyset, \{q\}\}$ , among which only the last two, i.e.,  $\{\emptyset, \{p\}\}$  and  $\{\emptyset, \{q\}\}$ , are the AEEMs of  $\varphi_2$  with respect to set inclusion. Note that  $\{\emptyset, \{p\}\} \approx_{\varphi_2} \{\emptyset, \{q\}\} \approx_{\varphi_2} \{\{p\}\} \approx_{\varphi_2} \{\{q\}\}$ . Finally, consider once again  $\varphi_3 = \neg K \neg p \rightarrow p$  of Table 4. We have  $\{\emptyset\} \models^* \varphi_3$  and  $\{\{p\}\} \models^* \varphi_3$ . We have seen above that  $\{\emptyset, \{p\}\} \models^* \varphi_3$ , while  $\{\emptyset, \{p\}\} \not\models^* \varphi_3$ . Therefore  $\{\{p\}\}$  is preferred over  $\{\emptyset\}$  with respect to  $\leq_{\varphi_3}$ , i.e., we have  $\{\emptyset\} \leq_{\varphi_3} \{\{p\}\}$  and  $\{\{p\}\} \not\leq_{\varphi_3} \{\emptyset\}$ . Moreover, since  $\{\{p\}\}$  and  $\{\emptyset\}$  are incomparable under set inclusion,  $\{\{p\}\}$  is the only AEEM of  $\varphi$ .

For Gelfond's 'eligibility' program  $\Pi_G$  of Section 3.1, the formula  $\Pi_G^*$  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  $\mathcal{T}_2 <_{\Pi_G^*} \mathcal{T}_1$ . So the only AEEM of  $\Pi_G$  is indeed  $\mathcal{T}_1 = \{\{h, e, i\}, \{f, i\}\}$ .

**Remark 3.** If  $\varphi$  has an EEM, then  $\varphi$  also has an AEEM. It follows from Proposition 7 that a nonepistemic formula  $\varphi$  has at most one AEEM: the set of all EMs of  $\varphi$ .

**Proposition 8.** If  $EEM(\varphi) = \emptyset$ , then  $\varphi$  has no AEEM. If  $EEM(\varphi)$  is a singleton, then the set of AEEMs of  $\varphi$  is equal to the set of EEMs of  $\varphi$ .

**Proof.** By definition, an AEEM of  $\varphi$  is an EEM of  $\varphi$  chosen according to two orderings. Then, the above result immediately follows.  $\square$

$\mathcal{T}$  is said to be an AEEM of an EHT theory  $\Phi$  if it is an AEEM of the conjunction of all formulas in  $\Phi$ . For example,  $\Phi = \{p, Kp \rightarrow (q \vee r)\}$  has three EEMs:  $\{\{p, q\}\}$ ,  $\{\{p, r\}\}$ , and  $\{\{p, q\}, \{p, r\}\}$ , among which only the last is an AEEM of  $\Phi$  with respect to set inclusion. Note that  $\{\{p, q\}\} \approx_{\Phi} \{\{p, r\}\} \approx_{\Phi} \{\{p, q\}, \{p, r\}\}$ .

## 6. AEEMs for epistemic specifications

We now compare AEEMs with the already-existing semantic approaches for epistemic specifications in the literature. The first thing to do is to translate epistemic specifications into EHT theories.

### 6.1. Translating epistemic specifications into EHT theories

Let  $\Pi$  be an epistemic specification. Our translation  $(\cdot)^*$  replaces ' $\leftarrow$ ', ' $\text{or}$ ', ' $\text{;}$ ' and ' $\text{not}$ ' respectively by ' $\rightarrow$ ', ' $\vee$ ', ' $\wedge$ ' and ' $\neg$ ', and the translation of  $K$  is direct. Simultaneously, we also invert the head and body parts of rules of  $\Pi$  and add parentheses when required to ensure a correct representation. Furthermore, we introduce a fresh variable  $\tilde{p}$  for each strongly-negated  $\sim p$  occurring in  $\Pi$ . For these new variables, the formula  $\text{Cons}(\Pi) = \bigwedge_{p \in \mathbb{P}_{\Pi}} \neg(p \wedge \tilde{p})$  guarantees 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  $\Pi$ .) Here is an example:

**Table 5**

Examples of epistemic specifications  $\Pi$  without disjunctions in the head (first column); Kahl et al.'s (latest version [7]) and Shen et al.'s world views (second column) of  $\Pi$ ; and our EEMs and AEEMs for the translations  $\Pi^*$  (third column; AEEMs in bold, together with all existing EEMs and the relations  $\subseteq$  or  $\leq_{\Pi^*}$  they are involved in).

Epistemic specification $\Pi$	World views	EEMs of $\Pi^*$
$p \leftarrow Kp$	$\{\emptyset\}$	$\{\emptyset\}$
$p \leftarrow \text{not } K\text{not } p$	$\{p\}$	$\{\emptyset\} <_{\Pi^*} \{\mathbf{p}\}$ (incomparable w.r.t. $\subseteq$ )
$p \leftarrow Kp$ $p \leftarrow \text{not } Kp$	none	<b>none</b>
$p \leftarrow \text{not } K\text{not } p$ $p \leftarrow K\text{not } p$	$\{p\}$	$\{\mathbf{p}\}$
$p \leftarrow Kq$ $q \leftarrow Kp$	$\{\emptyset\}$	$\{\emptyset\}$
$p \leftarrow \text{not } q$ $q \leftarrow \text{not } Kp$	$\{q\}$	$\{p\} <_{\Pi^*} \{\mathbf{q}\}$ (incomparable w.r.t. $\subseteq$ )
$p \leftarrow \text{not } Kq$ $q \leftarrow \text{not } Kp$	$\{p\}$ and $\{q\}$	$\{\mathbf{p}\}$ and $\{\mathbf{q}\}$ (incomparable w.r.t. $\leq_{\Pi^*}$ and $\subseteq$ )
$p \leftarrow q$ $q \leftarrow \text{not } K\text{not } p$	$\{p, q\}$	$\{\emptyset\} <_{\Pi^*} \{\mathbf{p}, \mathbf{q}\}$ (incomparable w.r.t. $\subseteq$ )
$p \leftarrow \text{not } K\text{not } q$ $q \leftarrow \text{not } K\text{not } p$	$\{p, q\}$	$\{\emptyset\} <_{\Pi^*} \{\mathbf{p}, \mathbf{q}\}$ (incomparable w.r.t. $\subseteq$ )
$p \leftarrow \text{not } K\text{not } q, \text{not } q$ $q \leftarrow \text{not } K\text{not } p, \text{not } p$	$\{p\}, \{q\}$	$\{\emptyset\} <_{\Pi^*} \{\mathbf{p}\}, \{\mathbf{q}\}$ (incomparable w.r.t. $\subseteq$ )

$$\Pi = \{p \text{ or } \sim q \leftarrow r, \text{not } s, q \leftarrow \text{not } K\text{not } p\};$$

$$\Pi^* = \{(r \wedge \neg s) \rightarrow (p \vee \bar{q})\} \wedge (\neg K\neg p \rightarrow q) \wedge \neg(q \wedge \bar{q}).$$

So,  $\text{Cons}(\Pi)$  corresponds to  $\neg(q \wedge \bar{q})$  in this example.

## 6.2. Comparison with Kahl et al.'s semantics

We now compare Kahl et al.'s world views of an epistemic specification  $\Pi$  with our AEEMs of  $\Pi^*$ . We do so by means of a series of examples most of which stem from [6].

Table 5 lists some simple epistemic specifications and their EEMs together with the relevant orderings; Table 6 contains slightly more complex examples with disjunctive heads. Let us examine the case of the epistemic specification

$$\Pi_0 = \{p \text{ or } q \leftarrow, p \leftarrow \text{not } Kq\}.$$

The EEMs of its translation  $\Pi_0^* = \{(p \vee q) \wedge (\neg Kq \rightarrow p)\}$  are  $\{p\}$  and  $\{q\}$ . It can be checked that  $\{\{p\}, \{q\}\} \models^* \Pi_0^*$  while  $\{\{p\}, \{q\}\} \not\models^* \Pi_0^*$ . Therefore  $\{p\} \not\leq_{\Pi_0^*} \{q\}$  and  $\{q\} \leq_{\Pi_0^*} \{p\}$ , which means that  $\{p\}$  is preferred over  $\{q\}$  with respect to  $\leq_{\Pi_0^*}$ .

Two critical examples of Table 5 that we would like to discuss here are  $\Pi_1 = \{p \leftarrow Kp, p \leftarrow \text{not } Kp\}$  and  $\Pi_2 = \{p \leftarrow \text{not } K\text{not } p, p \leftarrow K\text{not } p\}$ . Remember that  $Mp$  and  $\text{not } K\text{not } p$  are equivalent in Kahl et al.'s semantics, as well as  $K\text{not } p$  and  $\text{not } Mp$ ; so  $\Pi_2$  is equivalent to  $\{p \leftarrow Mp, p \leftarrow \text{not } Mp\}$ . Just as we, Kahl et al. obtain no world views for  $\Pi_1$  and a unique world view  $\{p\}$  for  $\Pi_2$ . Both results are clearly intuitive. However, for Kahl et al.'s semantics, the disjunctions of the two body parts of  $\Pi_1$  and  $\Pi_2$  are both tautologous, and  $\Pi_1$  and  $\Pi_2$  should therefore be both equivalent to  $\{p \leftarrow\}$ . For that reason, it is strange that  $\Pi_2$  has a solution while  $\Pi_1$  has none. In contrast, in our semantics  $Kp$  and  $\neg Kp$  are not equivalent; so  $Kp \vee \neg Kp$  is not tautologous, while  $\neg Kp \vee \neg \neg Kp$  is so, in accordance with the principles of intuitionistic logic. (Notice that the weak law of excluded middle  $\neg\varphi \vee \neg\neg\varphi$  is valid in EHT while the law of excluded middle  $\varphi \vee \neg\varphi$  is not.)

We note that for the last example of Table 5, the unique world view  $\{p\}, \{q\}$  of [7] coincides with our unique AEEM, while the original version of [6] had two world views, namely  $\{\emptyset\}$  and  $\{p\}, \{q\}$ , of which  $\{\emptyset\}$  conflicts with the intuitions about quantification over answer sets. Moreover, as mentioned before, Kahl had discussed a *preference* relation in his PhD thesis [6] for determining whether or not to believe different forms of extended literals. With respect to this order,  $Mp$  is easier to establish than  $p$  as it would require a lower degree of conviction. Similarly, a rational agent will accept  $\text{not } p$  by

**Table 6**

Examples of epistemic specifications  $\Pi$  with disjunctions in the head (first column); Kahl et al.'s (latest version [7]) and Shen et al.'s world views (second column) of  $\Pi$ ; and our EEMs and AEEMs for the translations  $\Pi^*$  (third column; AEEMs in bold, together with all existing EEMs and the relations  $\subseteq$  or  $\leq_{\Pi^*}$  they are involved in).

Epistemic specification $\Pi$	World views	EEMs of $\Pi^*$
$p \text{ or } q \leftarrow$	$\{\{p\}, \{q\}\}$	$\{\{p\}\} \subset \{\{\mathbf{p}\}, \{\mathbf{q}\}\}$ $\{\{q\}\} \subset \{\{\mathbf{p}\}, \{\mathbf{q}\}\}$ $\{\{p\}\} \approx_{\Pi^*} \{\{q\}\} \approx_{\Pi^*} \{\{p\}, \{q\}\}$
$p \text{ or } q \leftarrow$ $p \leftarrow \text{not } K \text{not } q$	$\{\{p\}\}$	$\{\{\mathbf{p}\}\}$
$p \text{ or } q \leftarrow$ $p \leftarrow Kq$	$\{\{p\}, \{q\}\}$	$\{\{p\}\} \subset \{\{\mathbf{p}\}, \{\mathbf{q}\}\}$ $\{\{p\}\} \approx_{\Pi^*} \{\{p\}, \{q\}\}$
$p \text{ or } q \leftarrow$ $p \leftarrow \text{not } Kq$	$\{\{p\}\}$	$\{\{q\}\} <_{\Pi^*} \{\{\mathbf{p}\}\}$ (incomparable w.r.t. $\subseteq$ )
$p \text{ or } q \leftarrow$ $r \leftarrow Kp$	$\{\{p\}, \{q\}\}$	$\{\{p, r\}\} <_{\Pi^*} \{\{\mathbf{p}\}, \{\mathbf{q}\}\}$ $\{\{q\}\} \subset \{\{\mathbf{p}\}, \{\mathbf{q}\}\}$ $\{\{q\}\} \approx_{\Pi^*} \{\{p\}, \{q\}\}$
$p \text{ or } q \leftarrow$ $r \leftarrow \text{not } K \text{not } p$	$\{\{p, r\}, \{q, r\}\}$	$\{\{p, r\}\} \subset \{\{\mathbf{p}, \mathbf{r}\}, \{\mathbf{q}, \mathbf{r}\}\}$ $\{\{q\}\} <_{\Pi^*} \{\{\mathbf{p}, \mathbf{r}\}, \{\mathbf{q}, \mathbf{r}\}\}$ $\{\{p, r\}\} \approx_{\Pi^*} \{\{p, r\}, \{q, r\}\}$
$p \leftarrow$ $q \text{ or } r \leftarrow Kp$	$\{\{p, q\}, \{p, r\}\}$	$\{\{p, q\}\} \subset \{\{\mathbf{p}, \mathbf{q}\}, \{\mathbf{p}, \mathbf{r}\}\}$ $\{\{p, r\}\} \subset \{\{\mathbf{p}, \mathbf{q}\}, \{\mathbf{p}, \mathbf{r}\}\}$ $\{\{p, q\}\} \approx_{\Pi^*} \{\{p, r\}\} \approx_{\Pi^*} \{\{p, q\}, \{p, r\}\}$

default over  $p$ . Kahl's discussion provides a stronger support for our result.<sup>10</sup> The first example of Table 6 where  $\{\{p\}\} \leq_{p \vee q} \{\{q\}\}$  and  $\{\{q\}\} \leq_{p \vee q} \{\{p\}\}$  illustrates that  $\leq_{\varphi}$  fails to be antisymmetric.

For all the examples in Tables 5 and 6, the last version of Kahl et al.'s semantics behaves exactly as ours. There are however differences between these two approaches. This can be illustrated by an example that is discussed in [14]. Consider the program

$$\{ p \leftarrow Mq, \text{not } q, q \leftarrow Mp, \text{not } p \}$$

for which  $\{\{p\}, \{q\}\}$  is the only world view per Kahl et al.'s knowledge minimisation. Adding  $r \leftarrow Mp, Mq$  to that program results in the world view  $\{\{p, r\}, \{q, r\}\}$ , as one might expect. But if one furthermore adds the rule  $s \leftarrow Kr$  then Kahl et al. [7] (and also Shen et al. [12]) obtain two world views  $\{\{p, r, s\}, \{q, r, s\}\}$  and  $\{\emptyset\}$  of which the latter is counterintuitive. In contrast, we only get the intended AEEM  $\{\{p, r, s\}, \{q, r, s\}\}$  for the whole program, the reason being that  $\{\{p, r, s\}, \{q, r, s\}\}$  is preferred over  $\{\emptyset\}$ .

### 6.3. Comparison with Shen et al.'s approach

In 2016, Shen et al. [12,20] proposed a new semantics for epistemic specifications. The idea is to use  $\text{not } K$  (which they call *epistemic negation*) to minimise knowledge in the set of all belief sets. Given an epistemic logic program  $\Pi$ , let  $\mathbb{E}_{\mathcal{P}}(\Pi)$  be the set of all epistemic negations in  $\Pi$  as defined in Section 3.2, and let  $\Phi \subseteq \mathbb{E}_{\mathcal{P}}(\Pi)$  be its subset, called a *guess*.

Given a non-empty collection  $\mathcal{A} \subseteq 2^{\mathcal{L}}$  of consistent sets of literals, let  $\Phi_{\mathcal{A}} = \{L \in \mathbb{E}_{\mathcal{P}}(\Pi) : \mathcal{A} \models L\} \subseteq \mathbb{E}_{\mathcal{P}}(\Pi)$  be the set of all epistemic negations in  $\Pi$  that are satisfied by  $\mathcal{A}$ . Then we transform  $\Pi$  into an epistemic reduct  $\Pi^{\Phi}$  with respect to  $\Phi$  by replacing every  $\text{not } K\lambda \in \Phi$  with  $\top$  and every  $\text{not } K\lambda \in \mathbb{E}_{\mathcal{P}}(\Pi) \setminus \Phi$  with  $\text{not } \lambda$  where  $\lambda$  is an extended objective literal. Finally, the collection  $\mathcal{A}$  is a world view of  $\Pi$  if

1.  $\mathcal{A} = \text{AS}(\Pi^{\Phi}) = \{A : A \text{ is an answer set of } \Pi^{\Phi}\}$ ;
2.  $\Phi_{\mathcal{A}}$  agrees with  $\Phi$ , i.e.,  $\Phi_{\mathcal{A}} = \Phi$ ;
3.  $\Phi$  is maximal, i.e., there is no bigger guess  $\Phi' \supset \Phi$  such that  $\mathcal{A}' = \text{AS}(\Pi^{\Phi'})$  and  $\Phi_{\mathcal{A}'} = \Phi'$  for some non-empty collection  $\mathcal{A}'$  of consistent sets of objective literals.

<sup>10</sup> Following this example and being inspired by the maximality condition mentioned in [12] (see Section 6.3, item 3), Kahl et al. came up with an update of [6] to address the issue, with semantics supporting only  $\{\{p\}, \{q\}\}$  (see [7,14]). The definition of world view of Section 3.2 identifies that maximality condition.



Let us illustrate their approach by the program of *closed world assumption* (CWA):

$$\Pi = \{\tilde{p} \leftarrow \text{not} K p, \perp \leftarrow p, \tilde{p}\}.$$

Then, take the guess  $\Phi = \{\text{not} K p\}$ . Thus,  $\Pi^\Phi = \{\tilde{p} \leftarrow \top, \perp \leftarrow p, \tilde{p}\}$ . Clearly,  $\text{AS}(\Pi^\Phi) = \{\{\tilde{p}\}\}$ , and  $\{\{\tilde{p}\}\} \models_{\text{ES}} \Phi$ . Since  $\Phi$  is the maximal guess possible (see item 3 above),  $\{\{\tilde{p}\}\}$  is the unique world view of  $\Pi$  according to Shen et al.'s semantics. Tables 5 and 6 contain more examples on this approach.

Shen et al. [12] argue that Pearce's equilibrium semantics suffers from circular justifications and that our approach inherits the same circularity. According to Shen et al., this leads to some undesired results in our approach. They illustrate their claim through two simple (epistemic) logic programs. First, they consider the program

$$\Pi_1 = \{p \leftarrow \text{not not } p, p \leftarrow \text{not } p\}.$$

The EHT theory  $\Pi_1^* = \{\neg\neg p \rightarrow p, \neg p \rightarrow p\}$  has a single EM  $\{p\}$  and therefore a unique AEEM  $\{\{p\}\}$ ; according to Shen et al.,  $p$  is justified via a self-supporting loop. However,<sup>11</sup> we know that already in intuitionistic logic, which is weaker than EHT and which incarnates the 'standard' of constructive reasoning,  $\neg p \rightarrow p$  is logically equivalent to  $\neg\neg p$ . Then  $p$  immediately follows from  $\neg\neg p \rightarrow p$  by Modus Ponens. From another perspective, the EHT theory  $\Pi_1^*$  amounts to the EHT formula  $(\neg\neg p \vee \neg p) \rightarrow p$ . Moreover, in HT and its epistemic extension EHT, the weak law of excluded middle  $\neg\neg p \vee \neg p$  is valid. Thus,  $\Pi_1^*$  is logically equivalent to  $p$ . As a result,  $\Pi_1$  should clearly have the world view  $\{\{p\}\}$ . The other example discussed by Shen et al. [12] is

$$\Pi_2 = \{p \leftarrow \text{not } K p \vee p\}.$$

The EHT theory  $\Pi_2^* = \{(\neg K p \vee p) \rightarrow p\}$  has no AEEMs, but [12] claims that  $\{\{p\}\}$  should be a world view of  $\Pi_2$ . They justify their result through the following argument: given a collection  $\mathcal{T}$  of valuations, since  $\mathcal{T}, T \models_{\text{ES}} \text{not } K p \vee p$  for every  $T \in \mathcal{T}$ , they assert  $\text{not } K p \vee p$  to be a tautology, and so they argue that  $\Pi_2$  amounts to  $p$ . However,  $\neg K p \vee p$  is not valid in EHT; one immediate countermodel is  $\{\{\emptyset, \{p\}\}\}$ . Thus, from a logical point of view,  $p$  does not follow from  $\Pi_2^*$ . From another perspective, we know that  $K p \rightarrow p$  is a theorem of EHT. We also note that contraposition holds for intuitionistic implication  $\rightarrow$  in EHT. Thus,  $\neg p \rightarrow \neg K p$  is also a theorem of EHT, but then so is  $\neg p \vee p \rightarrow \neg K p \vee p$ . As a result, we conclude that  $\neg K p \vee p$  is weaker than  $\neg p \vee p$  since any model of the latter is also a model of the former. Through a similar argument,  $\neg p \vee p \rightarrow p$  is weaker than  $\neg K p \vee p \rightarrow p$ , in other words, EHT models of the latter is a (strict) subset of those of the former. We know that  $\neg p \vee p \rightarrow p$  does not have an EM.<sup>12</sup> Then,  $\neg K p \vee p \rightarrow p$  cannot have an EEM either.

To sum it up, Shen et al.'s criticisms fail to provide evidence against our AEEMs.

#### 6.4. Comparison with older approaches

We here discuss some older approaches that tried to generalise or to refine epistemic specifications. All the approaches that we discuss here deal with the former versions [3–5] of epistemic specifications, which behave differently from the more recent versions [6,7,12]. We have already given an example in Table 4 distinguishing Gelfond's former and recent versions of epistemic specifications, which is also discussed in Section 3.2: for  $\Pi_1 = \{p \leftarrow K p\}$ , while the first version [3,4] gives two world views  $\{\emptyset\}$  and  $\{\{p\}\}$ , the second version [5] eliminates the unintended world view  $\{\{p\}\}$ . Moreover, the example  $\Pi_2 = \{p \leftarrow \text{not } K \text{not } p\}$  again in Table 4 shows that Kahl further improves Gelfond's recent version [5]: as Gelfond's approach offers two world views  $\{\emptyset\}$  and  $\{\{p\}\}$  for  $\Pi_2$ , Kahl's version [6] rejects the unintuitive one, viz.  $\{\emptyset\}$ . So, when it comes to having intuitive results especially for cyclic programs, our approach is closer to Kahl et al.'s [6,7] and Shen et al.'s [12] versions. However, we seriously depart from [12] in some basic examples, already mentioned in Section 6.3.

To the best of our knowledge, it was Chen who made the first attempt to embed epistemic specifications into a modal logic [8]. He was motivated by the close relation between the notion of *only knowing* [33] and the notion of *world view*. He proposed an epistemic modal logic with a kind of minimal model reasoning about epistemic concepts like knowledge and belief so as to capture epistemic specifications. However, it is fair to say that his logic has syntactically and semantically a complex nature due to the display of four different modal operators. As a result of this, it is a bit difficult to observe the intuition underlying it. Moreover, his approach significantly differs from ours because he embeds the first version of epistemic specifications [3,4] into a classical modal logic, yet our approach is in terms of an intuitionistic modal logic with non-dual epistemic modal operators and is closer to Kahl et al.'s recent version [7].

When it comes to structure, we are close, among all, to Wang and Zhang's (WZ) approach [9]. They described an epistemic extension of equilibrium logic into which they were able to embed Gelfond's first version [3,4] of epistemic specifications. This extension also gives semantics to nested epistemic logic programs. The language extends that of HT by

<sup>11</sup> We acknowledge a discussion with David Pearce on this issue, who put forward the argument below. Moreover, it is also explained in [34] in detail why equilibrium logic does not suffer those claimed circular justifications.

<sup>12</sup>  $\neg p \vee p \rightarrow p$  is logically equivalent to the HT theory  $\Gamma = \{\neg p \rightarrow p, p \rightarrow p\}$ . Since  $\neg p \rightarrow p$  is intuitionistically equivalent to  $\neg\neg p$ , we get that  $\Gamma$  is logically equivalent to  $\neg\neg p \wedge \top$ , and even further to  $\neg\neg p$ . We know that  $\neg\neg p$  does not have an EM. Thus,  $\neg p \vee p \rightarrow p$  cannot have an EM either.

two modal operators  $K$  and  $M$ . Let us call their epistemic HT models *WZ-EHT models*. A WZ-EHT model is a triple  $(\mathcal{A}, H, T)$  in which  $\mathcal{A} \subseteq 2^{\mathbb{P}}$  is a collection of valuations and  $(H, T)$  is an HT model such that  $H \subseteq T$ . Note that  $H$  and  $T$  are not necessarily contained in  $\mathcal{A}$ . For such  $\mathcal{A}$ , they define the collection of HT models

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

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

$$\begin{aligned} \mathcal{A}, H, T \models_{\text{EHT}} p & \quad \text{if } p \in H; \\ \mathcal{A}, H, T \models_{\text{EHT}} K\varphi & \quad \text{if } \text{coll}(\mathcal{A}), A \models_{\text{EHT}} \varphi \text{ for every } A \in \text{coll}(\mathcal{A}); \\ \mathcal{A}, H, T \models_{\text{EHT}} M\varphi & \quad \text{if } \text{coll}(\mathcal{A}), A \models_{\text{EHT}} \varphi \text{ for some } A \in \text{coll}(\mathcal{A}). \end{aligned}$$

Then, WZ-EEMs for a theory  $\Phi$  are total WZ-EHT models  $(\mathcal{A}, T, T)$  of  $\Phi$  such that there is no EHT model  $(\mathcal{A}, H, T)$  of  $\Phi$  with  $H \subset T$ . Finally, Wang and Zhang define equilibrium views by which they capture the world view notion of epistemic specifications. An *equilibrium view* of a theory  $\Phi$  is a maximal collection  $\mathcal{A} \subseteq 2^{\mathbb{P}}$  satisfying the fixed-point equation

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

Although both WZ's and our approaches propose an epistemic extension of equilibrium logic, there are some fundamental differences between them. Let us summarise the relationship.

1. HT can hardly be considered as a fragment of WZ-EHT because a WZ-EHT model  $(\mathcal{A}, H, T)$  has two isolated parts in which  $\mathcal{A}$  is concerned with the satisfiability of epistemic formulas and  $(H, T)$  is related to the satisfiability of non-epistemic formulas. As a result of this,  $Kp \wedge \neg p$  has WZ-EHT models such as  $(\{\{p\}, \{p, q\}\}, \emptyset, \{q\})$  and even WZ-EEMs such as  $(\{\{p\}\}, \emptyset, \emptyset)$  while it has none in our approach. Thus, each satisfiable EHT formula has also a WZ-EHT model; but not vice versa.
2. WZ-EHT models are much less in quantity than ours. For instance, WZ's approach can never generate collections like  $\{(\emptyset, \{p\})\}$  or  $\{(\emptyset, \emptyset), (\emptyset, \{p\})\}$  because the collection that corresponds to  $\mathcal{A} = \{\emptyset, \{p\}\}$  is precisely  $\{(\emptyset, \emptyset), (\emptyset, \{p\}), (\{p\}, \{p\})\}$ . Therefore, they cannot distinguish formulas like  $K\neg\neg p$  and  $Kp$ , nor  $M\neg\neg p$  and  $Mp$ , etc. Moreover, different from our approach, WZ-EHT is not an intuitionistic modal logic because the modal operators  $K$  and  $M$  are dual. As a result, our EHT version is more expressive than WZ-EHT.
3. There may be more WZ-EEMs than EEMs because the concept of minimality of truth differs: for instance, both  $Kp \wedge \neg p$  and  $K\neg\neg p$  have WZ-EEMs.  $(\{\{p, q, r\}\}, \emptyset, \emptyset)$  is a WZ-EEM for both of these formulas, but it is strange to see the propositional variables  $q$  and  $r$  in an equilibrium model, while they do not appear in the formulas. This way, we can put infinitely many variables into  $\mathcal{A} = \{\{p, q, r\}\}$ . As expected, these formulas have no EEMs in our sense. The other way around,  $Kp \wedge \neg\neg p$  has an EEM, viz.  $\{\{p\}\}$  in our sense, but no WZ-EEM.
4. Maximisation of ignorance is not performed in the same way:  $Kp$  has an AEEM, viz.  $\{\{p\}\}$ , but no WZ-equilibrium view (note that any WZ-EEM  $(\mathcal{A}, T, T)$  of  $Kp$  has  $T = \emptyset$ ). The formulas  $\hat{K}p$ ,  $\hat{K}p \wedge \neg p$ ,  $(p \vee q) \wedge (\hat{K}q \rightarrow p)$  and  $p \vee K(p \vee \neg p)$  are other examples; conversely,  $(\hat{K}\neg p \vee \hat{K}\neg\neg p) \wedge (\hat{K}p \rightarrow (p \vee \neg p))$  has a WZ-equilibrium view, but no AEEM.

More recently, Truszczyński [10] 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:  $(p \vee \neg p) \rightarrow p$  has no EMs, and therefore has no EEMs, nor world views according to Kahl's approach, while Shen et al. argue for the world view  $\{\{p\}\}$ .

## 7. 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 for any two EHT theories  $\Phi_1$  and  $\Phi_2$ , the EHT validity of the equivalence

$$K\left(\bigwedge \Phi_1\right) \leftrightarrow K\left(\bigwedge \Phi_2\right)$$

characterises that

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

Our proofs are non-trivial generalisations of Lifschitz et al.'s proof for HT logic [18]. They require the syntactical characterisation of EHT models. To that end, using Proposition 4, we may suppose without loss of generality that  $\mathbb{P}$  is *finite*. This allows us to describe possible worlds and EHT models by means of formulas.

## 7.1. Some characteristic formulas

We first define the following formulas: for an EHT model  $(\mathcal{T}, \mathfrak{h})$  and  $T \in 2^{\mathbb{P}}$ ,

$$\begin{aligned} \text{Total} &= \text{K} \bigwedge_{p \in \mathbb{P}} (p \vee \neg p) \\ \text{At}(T) &= \left( \bigwedge_{p \in T} p \right) \wedge \left( \bigwedge_{p \notin T} \neg p \right) \\ \text{Ch}(\mathfrak{h}) &= \text{K} \bigwedge_{T \in \mathcal{T}} \left( \neg \text{At}(T) \rightarrow \left( \left( \bigwedge_{p \in \mathfrak{h}(T)} p \right) \wedge \bigwedge_{p \in (T \setminus \mathfrak{h}(T))} (p \rightarrow \text{Total}) \right) \right). \end{aligned}$$

The formula Total captures that an EHT model is total. At( $T$ ) says that we are exactly at the there-world  $T$  (and not at a here-world  $H \subset T$ ). Given an S5 model  $\mathcal{T}$ , Ch( $\mathfrak{h}$ ) characterises a here-function  $h'$ : it says that  $h'$  equals either  $h$  or the identity function  $id$ .

**Lemma 1.** *Given a multipointed EHT model  $((\mathcal{T}, \mathfrak{h}), \mathcal{T}_0)$  and  $T_0, T \in \mathcal{T}$ , we have:*

1.  $(\mathcal{T}, \mathfrak{h}), \mathcal{T}_0 \models_{\text{EHT}} \text{Total}$  if and only if  $h = id$ ;
2.  $(\mathcal{T}, \mathfrak{h}), \mathcal{T}_0 \models_{\text{EHT}} \text{At}(T)$  if and only if  $T_0 = T = \mathfrak{h}(T_0)$ ;
3.  $(\mathcal{T}, \mathfrak{h}), \mathcal{T}_0 \models_{\text{EHT}} \neg \text{At}(T)$  if and only if  $T_0 = T$ ;
4.  $(\mathcal{T}, \mathfrak{h}), \mathcal{T}_0 \models_{\text{EHT}} \text{Ch}(\mathfrak{h})$  and  $\mathcal{T}, \mathcal{T}_0 \models_{\text{S5}} \text{Ch}(\mathfrak{h})$ ;
5. If  $(\mathcal{T}, h'), \mathcal{T}_0 \models_{\text{EHT}} \text{Ch}(h)$ , then  $h' = id$  or  $h' = h$ .

**Proof.** Let  $((\mathcal{T}, h), \mathcal{T}_0)$  be a multipointed EHT model, and let  $T_0, T \in \mathcal{T}$ .

1. Assume that  $(\mathcal{T}, \mathfrak{h}), \mathcal{T}_0 \models_{\text{EHT}} \text{Total}$ . So, we have  $(\mathcal{T}, \mathfrak{h}), \mathcal{T}_0 \models_{\text{EHT}} \text{K} \bigwedge_{p \in \mathbb{P}} (p \vee \neg p)$ , implying that  $(\mathcal{T}, \mathfrak{h}), T \models_{\text{EHT}} \bigwedge_{p \in \mathbb{P}} (p \vee \neg p)$  for every  $T \in \mathcal{T}$ . Thus,  $(\mathcal{T}, \mathfrak{h}), T \models_{\text{EHT}} p \vee \neg p$  for every  $p \in \mathbb{P}$  and every  $T \in \mathcal{T}$ . So we have:  $p \in h(T)$  or  $p \notin T$  for every  $p \in \mathbb{P}$  and every  $T \in \mathcal{T}$ . As a result,  $\mathfrak{h}(T) = T$  for every  $T \in \mathcal{T}$ , that is,  $\mathfrak{h} = id$ . Conversely, let  $\mathfrak{h} = id$ . Then,  $(\mathcal{T}, \mathfrak{h}), T \models_{\text{EHT}} p \vee \neg p$  for every  $p \in \mathcal{T}$  and every  $T \in \mathcal{T}$ , further implying that  $(\mathcal{T}, \mathfrak{h}), T \models_{\text{EHT}} \bigwedge_{p \in \mathbb{P}} (p \vee \neg p)$ , for every  $T \in \mathcal{T}$ . Thus,  $(\mathcal{T}, \mathfrak{h}), \mathcal{T}_0 \models_{\text{EHT}} \text{K} \bigwedge_{p \in \mathbb{P}} (p \vee \neg p)$ , that is,  $(\mathcal{T}, \mathfrak{h}), \mathcal{T}_0 \models_{\text{EHT}} \text{Total}$ .
2. Assume that  $(\mathcal{T}, \mathfrak{h}), \mathcal{T}_0 \models_{\text{EHT}} \text{At}(T)$ . So, we have  $(\mathcal{T}, \mathfrak{h}), \mathcal{T}_0 \models_{\text{EHT}} \left( \bigwedge_{p \in T} p \right) \wedge \left( \bigwedge_{p \notin T} \neg p \right)$ , implying that  $p \in \mathfrak{h}(T_0)$  for every  $p \in T$  and  $p \notin T_0$  for every  $p \notin T$ . Thus, we conclude that  $T \subseteq \mathfrak{h}(T_0) \subseteq T_0$  and  $T_0 \subseteq T$ . As a result,  $\mathfrak{h}(T_0) = T_0 = T$ . Conversely, let  $T_0 = T = \mathfrak{h}(T_0)$ . Clearly,  $\mathfrak{h}(T_0), T_0 \models_{\text{HT}} p$  for every  $p \in T$  and  $\mathfrak{h}(T_0), T_0 \models_{\text{HT}} \neg p$  for every  $p \notin T$ . Thus, we have  $\mathfrak{h}(T_0), T_0 \models_{\text{HT}} \left( \bigwedge_{p \in T} p \right) \wedge \left( \bigwedge_{p \notin T} \neg p \right)$ , that is,  $\mathfrak{h}(T_0), T_0 \models_{\text{HT}} \text{At}(T)$ . Note that At( $T$ ) is a non-epistemic EHT formula, so its satisfaction exclusively depends on designated worlds in an EHT model. Then it is easy to see that  $(\mathcal{T}, \mathfrak{h}), \mathcal{T}_0 \models_{\text{EHT}} \left( \bigwedge_{p \in T} p \right) \wedge \left( \bigwedge_{p \notin T} \neg p \right)$ , that is,  $(\mathcal{T}, \mathfrak{h}), \mathcal{T}_0 \models_{\text{EHT}} \text{At}(T)$ .
3. By Proposition 3.2, we have  $(\mathcal{T}, h), \mathcal{T}_0 \models_{\text{EHT}} \neg \text{At}(T)$  if and only if  $(\mathcal{T}, id), \mathcal{T}_0 \models_{\text{EHT}} \text{At}(T)$ . Then we take  $h = id$  in item 2, and the result immediately follows.
4. First, take  $\alpha_{T'} = \neg \text{At}(T') \rightarrow \left( \bigwedge_{p \in \mathfrak{h}(T')} p \wedge \bigwedge_{p \in (T' \setminus \mathfrak{h}(T'))} (p \rightarrow \text{Total}) \right)$ . By item 3, we have  $(\mathcal{T}, h), \mathcal{T}_0 \models_{\text{EHT}} \neg \text{At}(T)$  if and only if  $T = T_0$ . Thus, for an arbitrary  $T' \in \mathcal{T}$ ,  $(\mathcal{T}, \mathfrak{h}), T' \models_{\text{EHT}} \neg \text{At}(T')$ . Clearly,  $(\mathcal{T}, h), T' \models_{\text{EHT}} p$  for every  $p \in \mathfrak{h}(T')$  and  $(\mathcal{T}, \mathfrak{h}), T' \not\models_{\text{EHT}} p$  for every  $p \in (T' \setminus \mathfrak{h}(T'))$ ; further implying that  $(\mathcal{T}, h), T' \models_{\text{EHT}} \bigwedge_{p \in \mathfrak{h}(T')} p$  and  $(\mathcal{T}, h), T' \models_{\text{EHT}} p \rightarrow \text{Total}$  for every  $p \in (T' \setminus \mathfrak{h}(T'))$ . So,  $(\mathcal{T}, \mathfrak{h}), T' \models_{\text{EHT}} \left( \bigwedge_{p \in \mathfrak{h}(T')} p \right) \wedge \left( \bigwedge_{p \in (T' \setminus \mathfrak{h}(T'))} (p \rightarrow \text{Total}) \right)$ . Hence,  $(\mathcal{T}, \mathfrak{h}), T' \models_{\text{EHT}} \alpha_{T'}$ . On the other hand,  $(\mathcal{T}, h), T' \not\models_{\text{EHT}} \neg \text{At}(T)$  for any  $T \in \mathcal{T}$  such that  $T \neq T'$ . So,  $(\mathcal{T}, \mathfrak{h}), T' \models_{\text{EHT}} \alpha_T$  for any  $T \neq T'$ . As a result,  $(\mathcal{T}, \mathfrak{h}), T' \models_{\text{EHT}} \bigwedge_{T \in \mathcal{T}} \alpha_T$ . Since  $T' \in \mathcal{T}$  is arbitrary, we further get  $(\mathcal{T}, \mathfrak{h}), \mathcal{T}_0 \models_{\text{EHT}} \text{K} \bigwedge_{T \in \mathcal{T}} \alpha_T$ , that is,  $(\mathcal{T}, \mathfrak{h}), \mathcal{T}_0 \models_{\text{EHT}} \text{Ch}(h)$ . Moreover, it is easy to see that  $(\mathcal{T}, id), T' \models_{\text{EHT}} \alpha_T$  for any  $T \in \mathcal{T}$ . Note that  $(\mathcal{T}, id), T' \models_{\text{EHT}} \bigwedge_{p \in \mathfrak{h}(T')} p$ . Since our model is a total EHT model with an identity here function, we also have  $(\mathcal{T}, id), T' \models_{\text{EHT}} \bigwedge_{p \in (T' \setminus \mathfrak{h}(T'))} (p \rightarrow \text{Total})$ . Then, through the same reasoning as above, we get  $(\mathcal{T}, id), \mathcal{T}_0 \models_{\text{EHT}} \text{Ch}(h)$ , that is,  $\mathcal{T}, \mathcal{T}_0 \models_{\text{S5}} \text{Ch}(\mathfrak{h})$ .
5. Let  $(\mathcal{T}, h'), \mathcal{T}_0 \models_{\text{EHT}} \text{Ch}(h)$ . Without loss of generality, assume that  $h' \neq h$ . Then we need to show that  $h' = id$ . By assumption, since  $h' \neq \mathfrak{h}$ , there exists  $T' \in \mathcal{T}$  such that  $\mathfrak{h}(T') \neq h'(T')$  ( $\star$ ). Since Ch( $\mathfrak{h}$ ) starts with K, regardless of  $\mathcal{T}_0$ , we have  $(\mathcal{T}, \mathfrak{h}), T \models_{\text{EHT}} \text{Ch}(h)$  for every  $T \in \mathcal{T}$ . By item 3, we know that  $(\mathcal{T}, \mathfrak{h}'), T' \models_{\text{EHT}} \neg \text{At}(T')$ . Thus, we should also have  $(\mathcal{T}, \mathfrak{h}'), T' \models_{\text{EHT}} \bigwedge_{p \in \mathfrak{h}(T')} p$  and  $(\mathcal{T}, \mathfrak{h}'), T' \models_{\text{EHT}} \bigwedge_{p \in (T' \setminus \mathfrak{h}(T'))} (p \rightarrow \text{Total})$ . The former implies that  $h'(T'), T' \models_{\text{HT}} p$  for every  $p \in \mathfrak{h}(T')$ . Thus, we have  $\mathfrak{h}(T') \subseteq h'(T')$ . Since  $\mathfrak{h}(T') \neq h'(T')$  by ( $\star$ ), there exists  $p' \in h'(T') \setminus \mathfrak{h}(T')$ . As  $(\mathcal{T}, \mathfrak{h}'), T' \models_{\text{EHT}} \bigwedge_{p \in (T' \setminus \mathfrak{h}(T'))} (p \rightarrow \text{Total})$  and  $(\mathcal{T}, \mathfrak{h}'), T' \models_{\text{EHT}} p'$ , we have  $(\mathcal{T}, \mathfrak{h}'), T' \models_{\text{EHT}} \text{Total}$ . Then, by item 1,  $h' = id$ .  $\square$

## 7.2. Strong equivalence for EEMs

**Lemma 2.** Let  $((\mathcal{T}, \mathfrak{h}), \mathcal{T}_0)$  be a multipointed EHT model such that  $\mathfrak{h}$  is the identity on  $\mathcal{T} \setminus \mathcal{T}_0$ . Then,  $(\mathcal{T}, \mathfrak{h}), \mathcal{T}_0 \models_{\text{EHT}} \Phi_1$  and  $(\mathcal{T}, \mathfrak{h}), \mathcal{T}_0 \not\models_{\text{EHT}} \Phi_2$  implies that there is an EHT theory  $\Theta$  such that

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

**Proof.** We build two different  $\Theta$ , depending on whether  $\mathcal{T}, \mathcal{T}_0 \models_{\text{SS}} \Phi_2$  or not.

Case 1: let  $\mathcal{T}, \mathcal{T}_0 \not\models_{\text{SS}} \Phi_2$ . Then, we define  $\Theta = \{\text{Total}\}$  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 follows from the hypothesis that  $\mathcal{T}, \mathcal{T}_0 \not\models_{\text{SS}} \Phi_2$ . As to the former,  $\mathcal{T}, \mathcal{T}_0 \models_{\text{SS}} \Phi_1 \cup \Theta$  is the case: (i) using the assumption that  $(\mathcal{T}, \mathfrak{h}), \mathcal{T}_0 \models_{\text{EHT}} \Phi_1$  and by Proposition 2, we have  $(\mathcal{T}, \text{id}), \mathcal{T}_0 \models_{\text{EHT}} \Phi_1$ , i.e.,  $\mathcal{T}, \mathcal{T}_0 \models_{\text{SS}} \Phi_1$ . (ii) By Lemma 1.1, we also have  $\mathcal{T}, \mathcal{T}_0 \models_{\text{SS}} \text{Total}$  since  $\mathfrak{h} = \text{id}$ . Thus,  $\mathcal{T}, \mathcal{T}_0 \models_{\text{SS}} \Phi_1 \cup \{\text{Total}\}$ . Moreover, again by Lemma 1.1, for every  $h' \neq \text{id}$ , we have  $(\mathcal{T}, h'), \mathcal{T}_0 \not\models_{\text{EHT}} \text{Total}$ ; in particular,  $(\mathcal{T}, h'), \mathcal{T}_0 \not\models_{\text{EHT}} \text{Total}$  for every  $h' \neq \text{id}$  such that  $h'|_{\mathcal{T} \setminus \mathcal{T}_0} = \text{id}$ . Hence, for every  $h' \neq \text{id}$  such that  $h'|_{\mathcal{T} \setminus \mathcal{T}_0} = \text{id}$ , we have  $(\mathcal{T}, h'), \mathcal{T}_0 \not\models_{\text{EHT}} \Phi_1 \cup \{\text{Total}\}$ . As a result,  $\mathcal{T}, \mathcal{T}_0 \models^* \Phi_1 \cup \Theta$ .

Case 2: let  $\mathcal{T}, \mathcal{T}_0 \models_{\text{SS}} \Phi_2$ . Then, we define  $\Theta = \{\text{Ch}(h)\}$ . We want to 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: (i) by our initial assumption we have  $(\mathcal{T}, h), \mathcal{T}_0 \models_{\text{EHT}} \Phi_1$  and by Lemma 1.4,  $(\mathcal{T}, h), \mathcal{T}_0 \models_{\text{EHT}} \text{Ch}(h)$ . Thus,  $(\mathcal{T}, \mathfrak{h}), \mathcal{T}_0 \models_{\text{EHT}} \Phi_1 \cup \Theta$  for some  $h$ ; (ii) while the function  $h$  is identity on  $\mathcal{T} \setminus \mathcal{T}_0$  by hypothesis, it is different from the identity function on the whole  $\mathcal{T}$  (note that  $\mathcal{T}, \mathcal{T}_0 \models_{\text{SS}} \Phi_2$  and  $(\mathcal{T}, h), \mathcal{T}_0 \not\models_{\text{EHT}} \Phi_2$ ). Consequently,  $\mathcal{T}, \mathcal{T}_0 \not\models^* \Phi_1 \cup \Theta$ . As to the latter (i.e.,  $\mathcal{T}, \mathcal{T}_0 \models^* \Phi_2 \cup \Theta$ ): first, we have  $\mathcal{T}, \mathcal{T}_0 \models_{\text{SS}} \Phi_2$  by hypothesis and  $\mathcal{T}, \mathcal{T}_0 \models_{\text{SS}} \text{Ch}(h)$  by Lemma 1.4. Thus,  $\mathcal{T}, \mathcal{T}_0 \models_{\text{SS}} \Phi_2 \cup \Theta$ . Then, we consider some  $h' \neq \text{id}$  on  $\mathcal{T}$  such that  $h'|_{\mathcal{T} \setminus \mathcal{T}_0} = \text{id}$ . On the one hand, let  $h' = h$  (remember that  $h \neq \text{id}$  on  $\mathcal{T}$ ), then by our initial assumption,  $(\mathcal{T}, h'), \mathcal{T}_0 \not\models_{\text{EHT}} \Phi_2$ ; on the other hand, let  $h' \neq h$  then since  $h' \neq \text{id}$ , by Lemma 1.5,  $(\mathcal{T}, h'), \mathcal{T}_0 \not\models_{\text{EHT}} \text{Ch}(h)$ . As a result,  $(\mathcal{T}, h'), \mathcal{T}_0 \not\models_{\text{EHT}} \Phi_2 \cup \{\text{Ch}(h)\}$  for every  $h' \neq \text{id}$  such that  $h'|_{\mathcal{T} \setminus \mathcal{T}_0} = \text{id}$ . Thus,  $\mathcal{T}, \mathcal{T}_0 \not\models^* \Phi_2 \cup \Theta$ .  $\square$

**Theorem 1.** Let  $\Phi_1$  and  $\Phi_2$  be two EHT theories. Then the following are equivalent:

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

**Proof.** For the right-to-left direction, suppose  $K(\bigwedge \Phi_1) \leftrightarrow K(\bigwedge \Phi_2)$  is EHT valid, that is, no EHT model allows to distinguish  $K(\bigwedge \Phi_1)$  and  $K(\bigwedge \Phi_2)$ . Then, by Proposition 1, we have  $(\mathcal{T}, \mathfrak{h}), \mathcal{T} \models_{\text{EHT}} \Phi_1$  if and only if  $(\mathcal{T}, \mathfrak{h}), \mathcal{T} \models_{\text{EHT}} \Phi_2$ , for every EHT model  $(\mathcal{T}, h)$ . Therefore, for every EHT theory  $\Theta$ ,  $(\mathcal{T}, \mathfrak{h}), \mathcal{T} \models_{\text{EHT}} \Phi_1 \cup \Theta$  if and only if  $(\mathcal{T}, \mathfrak{h}), \mathcal{T} \models_{\text{EHT}} \Phi_2 \cup \Theta$ , for every EHT model  $(\mathcal{T}, h)$ . Thus, it follows that  $EEM(\Phi_1 \cup \Theta) = EEM(\Phi_2 \cup \Theta)$ , for every  $\Theta$ .

For the left-to-right direction, suppose that the equivalence  $K(\bigwedge \Phi_1) \leftrightarrow K(\bigwedge \Phi_2)$  is not EHT valid. Without loss of generality, suppose that  $(\mathcal{T}, \mathfrak{h}), \mathcal{T}_0 \models_{\text{EHT}} K(\bigwedge \Phi_1)$  and  $(\mathcal{T}, \mathfrak{h}), \mathcal{T}_0 \not\models_{\text{EHT}} K(\bigwedge \Phi_2)$ , for some EHT model  $(\mathcal{T}, \mathfrak{h})$  and  $\mathcal{T}_0 \in \mathcal{T}$ . Then, by Proposition 1, we have  $(\mathcal{T}, \mathfrak{h}), \mathcal{T} \models_{\text{EHT}} \Phi_1$  and  $(\mathcal{T}, \mathfrak{h}), \mathcal{T} \not\models_{\text{EHT}} \Phi_2$ . By Lemma 2 (which is applicable because  $\mathfrak{h}$  is trivially identity on  $\mathcal{T} \setminus \mathcal{T} = \emptyset$ ), there exists an EHT theory  $\Theta$  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  $\mathcal{T}, \mathcal{T} \not\models^* \Phi_2 \cup \Theta$ . As we have noted in Section 5.3 (see Remark 2), we know that  $\mathcal{T}, \mathcal{T} \models^* \Gamma$  if and only if  $\mathcal{T} \in EEM(\Gamma)$ , for any EHT theory  $\Gamma$ . So,  $EEM(\Phi_1 \cup \Theta) \neq EEM(\Phi_2 \cup \Theta)$  for some  $\Theta$ .  $\square$

## 7.3. Strong equivalence for AEEMs

We now define another formula characterising the set of all there-worlds of an EHT model  $(\mathcal{T}, \mathfrak{h})$  by means of Jankov-Fine-like formulas [35]:

$$\text{JF}(\mathcal{T}) = \left( \bigwedge_{T \in \mathcal{T}} \hat{K} \neg \neg \text{At}(T) \right) \wedge K \left( \bigvee_{T \in \mathcal{T}} \neg \neg \text{At}(T) \right).$$

**Lemma 3.** Let  $(\mathcal{S}, \mathfrak{h})$  be an EHT model. Then

$$(\mathcal{S}, \mathfrak{h}), \mathcal{S}_0 \models_{\text{EHT}} \text{JF}(\mathcal{T}) \text{ if and only if } \mathcal{S} = \mathcal{T}.$$

**Proof.** For the left-to-right direction, suppose that  $(\mathcal{S}, \mathfrak{h}), \mathcal{S}_0 \models_{\text{EHT}} \text{JF}(\mathcal{T})$ . Thus, we have  $(\mathcal{S}, \mathfrak{h}), \mathcal{S}_0 \models_{\text{EHT}} \bigwedge_{T \in \mathcal{T}} \hat{K} \neg \neg \text{At}(T)$  and  $(\mathcal{S}, \mathfrak{h}), \mathcal{S}_0 \models_{\text{EHT}} K \left( \bigvee_{T \in \mathcal{T}} \neg \neg \text{At}(T) \right)$ . By Lemma 1.3, the former implies that for every  $T \in \mathcal{T}$  there is  $S \in \mathcal{S}$  such that  $T = S$ . Thus,  $\mathcal{T} \subseteq \mathcal{S}$ . The latter implies that for every  $S \in \mathcal{S}$ , there exists  $T \in \mathcal{T}$  such that  $S = T$ . Hence,  $\mathcal{S} \subseteq \mathcal{T}$ . As a result,  $\mathcal{S} = \mathcal{T}$ .

For the right-to-left direction, let  $\mathcal{S} = \mathcal{T}$ . Then, we need to prove that  $(\mathcal{T}, \mathfrak{h}), \mathcal{T}_0 \models_{\text{EHT}} \text{JF}(\mathcal{T})$ . By Lemma 1.3,  $(\mathcal{T}, \mathfrak{h}), T \models_{\text{EHT}} \neg\neg\text{At}(T)$  for every  $T \in \mathcal{T}$ . Since the satisfiability of  $\hat{K}$  is independent of the set of designated worlds, we have  $(\mathcal{T}, \mathfrak{h}), \mathcal{T}_0 \models_{\text{EHT}} \bigwedge_{T \in \mathcal{T}} \hat{K} \neg\neg\text{At}(T)$ . Again by Lemma 1.3, since  $(\mathcal{T}, \mathfrak{h}), T \models_{\text{EHT}} \neg\neg\text{At}(T)$  for every  $T \in \mathcal{T}$ , we moreover have  $(\mathcal{T}, \mathfrak{h}), \mathcal{T} \models_{\text{EHT}} \bigvee_{T \in \mathcal{T}} \neg\neg\text{At}(T)$ . Then, by Proposition 1,  $(\mathcal{T}, \mathfrak{h}), \mathcal{T}_0 \models_{\text{EHT}} K(\bigvee_{T \in \mathcal{T}} \neg\neg\text{At}(T))$ . As a result,  $(\mathcal{T}, \mathfrak{h}), \mathcal{T}_0 \models_{\text{EHT}} \text{JF}(\mathcal{T})$ .  $\square$

**Theorem 2.** *Let  $\Phi_1$  and  $\Phi_2$  be two EHT theories. Then the following are equivalent:*

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

**Proof.** The proof of the right-to-left direction follows the lines of that of Theorem 1 (recall that when  $\text{EEM}(\Phi_1 \cup \Theta) = \text{EEM}(\Phi_2 \cup \Theta)$  for every EHT theory  $\Theta$ , it remains to check whether or not we choose the same EEMs as AEEMs with respect to set inclusion and preference ordering for the formulas  $\Phi_1 \cup \Theta$  and  $\Phi_2 \cup \Theta$ . Selection with respect to set inclusion is not related to the formulas, so it is obvious. As to the preference ordering, this selection is related to the formulas, but since  $\Phi_1 \cup \Theta$  and  $\Phi_2 \cup \Theta$  have the same Eht models, it is easy to see by definition that we again select the same EEMs with respect to this ordering as well. Thus,  $\Phi_1 \cup \Theta$  and  $\Phi_2 \cup \Theta$  have the same AEEMs).

The proof of the left-to-right direction also follows the lines of that of Theorem 1. (It is only the construction of the EHT theory  $\Theta$  of Lemma 2 that has to be adapted to this proof: the Jankov-Fine formula  $\text{JF}(\mathcal{T})$  has to be conjoined with the  $\Theta$ 's in the proof of Lemma 2.) Assume that  $K(\bigwedge \Phi_1) \leftrightarrow K(\bigwedge \Phi_2)$  is not EHT valid. Then there exists an EHT model  $(\mathcal{T}, \mathfrak{h})$  such that  $(\mathcal{T}, \mathfrak{h}), \mathcal{T} \models_{\text{EHT}} \Phi_1$  and  $(\mathcal{T}, \mathfrak{h}), \mathcal{T} \not\models_{\text{EHT}} \Phi_2$ , or just the opposite. Without loss of generality, assume that  $(\mathcal{T}, \mathfrak{h}), \mathcal{T} \models_{\text{EHT}} \Phi_1$  and  $(\mathcal{T}, \mathfrak{h}), \mathcal{T} \not\models_{\text{EHT}} \Phi_2$ . Then, by Lemma 2, there is an EHT theory  $\Theta$  in which  $\text{JF}(\mathcal{T})$  is contained such that  $\mathcal{T}, \mathcal{T} \models^* \Phi_1 \cup \Theta$  and  $\mathcal{T}, \mathcal{T} \not\models^* \Phi_2 \cup \Theta$  or vice versa. Lemma 3 guarantees that  $(\mathcal{T}, \text{id}), \mathcal{T} \models_{\text{EHT}} \text{JF}(\mathcal{T})$  and for every  $\mathcal{T}' \subseteq 2^{\mathbb{P}}$  such that  $\mathcal{T}' \neq \mathcal{T}$ ,  $(\mathcal{T}', \text{id}), \mathcal{T}' \not\models_{\text{EHT}} \text{JF}(\mathcal{T})$ . Thus, adding  $\text{JF}(\mathcal{T})$  into the construction of  $\Theta$  makes the result obtained in Theorem 1 unique for  $\mathcal{T}$ , i.e., for such  $\Theta$  in which  $\text{JF}(\mathcal{T})$  is added, there is exactly one EEM of  $\Phi_1 \cup \Theta$  (which is  $\mathcal{T}$  itself) and  $\Phi_2 \cup \Theta$  has none, or the other way around. It then follows from Proposition 8 that there is exactly one AEEM of  $\Phi_1 \cup \Theta$ , but  $\Phi_2 \cup \Theta$  has none, or vice versa. As a result, for such  $\Theta$ , the AEEMs of  $\Phi_1 \cup \Theta$  and  $\Phi_2 \cup \Theta$  must be different.  $\square$

As a direct corollary of Theorem 1 and Theorem 2, we have the result below:

**Corollary 3.** *Let  $\Phi_1$  and  $\Phi_2$  be two EHT theories. Then the following are equivalent:*

- (a)  $K(\bigwedge \Phi_1) \leftrightarrow K(\bigwedge \Phi_2)$  is EHT valid;
- (b)  $\Phi_1$  and  $\Phi_2$  are strongly equivalent in the sense of EEMs;
- (c)  $\Phi_1$  and  $\Phi_2$  are strongly equivalent in the sense of AEEMs;

As pointed out by one of our reviewers, this fact can also be proved in a simpler way by following the procedure: (a) immediately implies (b) since  $K(\bigwedge \Phi_1)$  and  $K(\bigwedge \Phi_2)$  have the same EHT models (note that EEMs are total EHT models satisfying a truth-minimising condition); (b) implies (c) because AEEMs are special forms of EEMs satisfying some knowledge-minimising conditions; (c) implies (a) by the second paragraph of the proof of Theorem 2 given above.

## 8. Semantics of EHT revisited: functional versus relational

Speaking roughly, we have said in Section 4 that our EHT models are basically sets of HT models, but this is not exactly the case. For example, the set

$$\{(\{p\}, \{p, q\}), (\{q\}, \{p, q\})\}$$

has no EHT model directly matching it because it allows multiple here worlds (i.e.,  $\{p\}$  and  $\{q\}$ ) for the same there world  $\{p, q\}$ . In this section, we introduce an alternative semantics for EHT allowing such situations, but also show that the two semantics however lead to the same logic: they give precisely the same epistemic and autoepistemic equilibrium models of a formula up to bisimilarity. Our original semantics is in functional style and should be a more natural presentation for readers that are familiar with intuitionistic modal logics. The advantage of the presentation in functional style of Section 4 is that it provides less and more concise models and is therefore simpler to work with.

From now on, we will call EHT models in the form of  $(\mathcal{T}, \mathfrak{h})$  *functional EHT models*, and accordingly call the original semantics of EHT *functional semantics* as defined in the previous sections. The new semantics given below is in relational style. It provides a nice modal logic perspective to EHT models.

A *relational EHT model* is a triple  $(W, V^H, V^T)$  in which

- $W$  is a non-empty set of possible worlds;

- $V^H, V^T : W \rightarrow 2^{\mathbb{P}}$  are functions assigning to each world  $w \in W$  respectively valuations  $V^H(w)$  and  $V^T(w)$  such that  $V^H(w) \subseteq V^T(w) \subseteq \mathbb{P}$ .

We call  $V^H$  a *here-valuation* and  $V^T$  a *there-valuation*. When  $V^H = V^T$ , we get a total relational EHT model which is only the other representation of a classical S5 model.

The relational version has a similar definition of bisimilarity: given two relational EHT models  $(W_1, V_1^H, V_1^T)$  and  $(W_2, V_2^H, V_2^T)$ , and  $P \subseteq \mathbb{P}$ , we call a relation  $Z \subseteq W_1 \times W_2$  *P-bisimulation* if

1. both  $Z$  and  $Z^{-1}$  are serial;
2. if  $w_1 Z w_2$  then  $V_1^H(w_1) \cap P = V_2^H(w_2) \cap P$  and  $V_1^T(w_1) \cap P = V_2^T(w_2) \cap P$ .

If there exists a *P-bisimulation*  $Z$  between  $(W_1, V_1^H, V_1^T)$  and  $(W_2, V_2^H, V_2^T)$  such that  $w_1 Z w_2$  then the pointed models  $((W_1, V_1^H, V_1^T), w_1)$  and  $((W_2, V_2^H, V_2^T), w_2)$  are called *P-bisimilar*.

Similar to a functional EHT model  $(\mathcal{T}, \mathfrak{h})$ , a relational EHT model  $(W, V^H, V^T)$  also determines a collection  $\{(V^H(w), V^T(w))\}_{w \in W}$  of HT models. This collection is now a multiset, and so allows for HT models  $(H_1, T_1)$  and  $(H_2, T_2)$  in which  $T_1 = T_2$ . As a result, as well as being able to describe all models in functional style, the relational model definition supplies many more additional models: some nontrivial new models such as  $\{(\emptyset, \{p\}), (\{p\}, \{p\})\}$  and some trivial models such as  $\{(\emptyset, \{p\}), (\emptyset, \{p\})\}$  and  $\{(\emptyset, \{q\}), (\{p\}, \{p\}), (\{p\}, \{p\})\}$  which are bisimilar to already-existing functional EHT models  $\{(\emptyset, \{p\})\}$  and  $\{(\emptyset, \{q\}), (\{p\}, \{p\})\}$  respectively. For example, if we restrict  $\mathbb{P}$  to  $\{p, q\}$ , then the formula  $\hat{K}(p \wedge \neg q) \wedge \hat{K}(q \wedge \neg p)$  has a unique (total) model  $\{(\{p, q\}, \{p, q\})\}$  in functional style whereas it has an additional model in relational style:  $\{(\{p\}, \{p, q\}), (\{q\}, \{p, q\})\}$ . Note that the latter model helps us distinguish  $p$  and  $\neg p$ .

A nonepistemic HT model  $(H, T)$  can be represented in the relational style as  $(W, V^H, V^T)$  where  $V^H(w) = H$  and  $V^T(w) = T$  for every  $w \in W$ . For a pointed relational EHT model, truth conditions and EHT validity are defined in a natural way.

Given two relational EHT models  $(W_1, V_1^H, V_1^T)$  and  $(W_2, V_2^H, V_2^T)$ , we write

$$(W_1, V_1^H, V_1^T) \trianglelefteq (W_2, V_2^H, V_2^T)$$

if  $W_1 = W_2$ ,  $V_1^T = V_2^T$  and  $V_1^H(w) \subseteq V_2^H(w)$ , for every  $w \in W_1$ , and we say that  $(W_1, V_1^H, V_1^T)$  is *weaker* than  $(W_2, V_2^H, V_2^T)$ . The strict version is defined in the usual way. For example,  $\{(\emptyset, \{p\}), (\{p\}, \{p\})\} \triangleleft \{(\{p\}, \{p\}), (\{p\}, \{p\})\}$  in the relational style.

Epistemic equilibrium models (EEMs) of EHT in the relational style are defined in the standard way:  $(W, V^T, V^T)$  is an EEM of  $\varphi$  if

1.  $(W, V^T, V^T), w \models_{\text{ss}} \varphi$  for every  $w \in W$ ;
2.  $(W, V^H, V^T), w \not\models_{\text{EHT}} \varphi$  for some  $w \in W$  and for every function  $V^H : W \rightarrow 2^{\mathbb{P}}$  such that  $(W, V^H, V^T) \triangleleft (W, V^T, V^T)$ .

**Proposition 9.** *The EEMs of a formula  $\varphi$  in the functional style respectively the relational style are equivalent up to bisimilarity.*

**Proof.**

- Let  $\mathcal{T} \subseteq 2^{\mathbb{P}}$  be an EEM of  $\varphi$  in the functional EHT semantics. Take  $W = \mathcal{T}$ ,  $V^T = id$ . Then the result follows.
- Let  $(W, V^T)$  be an EEM of  $\varphi$  over the relational EHT semantics  $(\star)$ . Suppose that  $(\mathcal{T}, id)$  is the bisimulation contraction of  $(W, V^T)$ . Hence, using the assumption  $(\star)$ , we have  $\mathcal{T}, \mathcal{T} \models_{\text{ss}} \varphi$  (•). Again by  $(\star)$ , we also know that for every  $(W, V^H, V^T)$  with  $(W, V^H, V^T) \triangleleft (W, V^T, V^T)$ , we have  $(W, V^H, V^T), w \not\models_{\text{EHT}} \varphi$  for some  $w \in W$ . For every such  $V^H$ , there is a function  $\mathfrak{h} : \mathcal{T} \rightarrow 2^{\mathbb{P}}$  such that  $(\mathcal{T}, \mathfrak{h})$  is a bisimulation contraction of  $(W, V^H, V^T)$ , and so  $(\mathcal{T}, \mathfrak{h}), T \not\models_{\text{EHT}} \varphi$  (••) in which  $T = V^T(w)$ . Hence, from (•) and (••) we obtain that  $\mathcal{T}$  is also an EEM of  $\varphi$ .  $\square$

It follows as a corollary of Proposition 9 that the autoepistemic equilibrium models (AEEMs) of a formula  $\varphi$  are also equivalent in both the functional and the relational presentations.

## 9. Discussion

We have already discussed older semantics proposals in Section 6.2, Section 6.3 and Section 6.4. In this section, we discuss a new and more principled line of research about epistemic specifications. Cabalar et al. [23,36] have very recently introduced some fundamental principles of knowledge modelling that should be used to evaluate the semantics proposals for epistemic specifications. Thus, according to their viewpoint, a correct semantics of epistemic specifications should be compatible with these foundational properties.

To begin with, we introduce a formal property proposed and named *epistemic splitting* by Cabalar et al. [23]. This property allows for a kind of modularity that, as the authors claim, guarantees a reasonable behaviour of programs whose subjective

literals are stratified. The idea is to separate a program  $\Pi$  into two disjoint subprograms (if possible), *top* and *bottom*, such that the subprogram top queries bottom through its subjective literals, and the subprogram bottom never refers to the head literals of top. If splitting is the case with respect to a set  $U$  of literals (called *splitting set*), then we calculate the world views of  $\Pi$  in four steps: first, we compute the world views  $W_b$  of bottom; second, for each  $W_b$ , we take a kind of partial reduct  $\Pi_U^{W_b}$  by replacing the subjective literals (whose literals are included in  $U$ ) of top with their truth values in  $W_b$ ; third, we find the world views  $W_t$  of  $\Pi_U^{W_b}$  and end with a solution  $(W_b, W_t)$  for  $\Pi$ ; finally, we concatenate the components of  $(W_b, W_t)$  in a specific way, resulting in the world views of the original program  $\Pi$ .

As already pointed out in [23], all proposed semantics of epistemic specifications but Gelfond's first version [3] fail to satisfy epistemic splitting property. In particular, Gelfond's subsequent refined versions do not satisfy it either. In [23], two counterexamples are given demonstrating this general fact. However, Gelfond's first approach suffers from the counterintuitive behaviour of cyclic programs more than all other approaches.<sup>13</sup> The other semantics that passes the epistemic splitting test is Truszczyński's approach [10].<sup>14</sup> Very recently, Cabalar et al. [36] have proposed a new epistemic extension of equilibrium logic called *founded autoepistemic equilibrium logic*, and Fandino [37] has proved that it also satisfies epistemic splitting.

We now present another property, which is closely associated with epistemic splitting (see Property 5 in [23]). This property regulates the functioning of *subjective constraints*, i.e., constraints whose body is composed of only the conjuncts of extended subjective literals. Formally, a semantics satisfies *subjective constraint monotonicity* if, for any epistemic logic program  $\Pi$  and any subjective constraint  $r$ ,  $W$  is a world view of  $\Pi \cup \{r\}$  if and only if both  $W$  is a world view of  $\Pi$  and  $W \models r$ . In other words, a constraint of this kind at most rules out the previously obtained world views of a program, but never generates new solutions. The AEEM semantics, however, does not satisfy this property either. As an immediate counterexample, let us consider the epistemic logic program  $\Pi = \{p \text{ or } q \leftarrow, \leftarrow \text{ not } K p\}$ . We have seen that  $p \vee q$  has a unique AEEM model  $\{\{p\}, \{q\}\}$ . When we consider the EHT theory  $\Pi^* = \{p \vee q, \neg \leftarrow K p\}$ , we see that it has a unique AEEM  $\{\{p\}\}$ . Notice that we would have had no AEEMs if it had satisfied the property of subjective constraint monotonicity:  $\{\{p\}, \{q\}\} \not\models \neg \leftarrow K p$ , so the EHT formula  $\neg \leftarrow K p$  corresponding to the subjective constraint  $\leftarrow \text{ not } K p$  would eliminate the AEEM  $\{\{p\}, \{q\}\}$  and eventually we would have no AEEMs at all.

Among the properties defined by Cabalar et al. there are also *supra-ASP* (the unique world view of a nonepistemic ASP program is the set of all its answer sets if they exist), *supra-S5* (any world view of an epistemic logic program is an S5 model), and a kind of derivability property called *foundedness* ensuring that self-supported world views of a program are rejected.

To sum up, researchers have not agreed yet on a fully satisfactory semantics for epistemic specifications, so the subject is still under progress, with several other recent publications [38–41].

## 10. Conclusion

In this paper, we have designed the logic of epistemic here-and-there (EHT): we have added two modal operators  $K$  and  $\hat{K}$  to the original language of HT logic and interpreted them in models that are the natural combinations of the HT models of here-and-there logic and the Kripke models for S5. We have shown that EHT is a particular intuitionistic modal logic where  $K$  and  $\hat{K}$  are not dual. Following this, we have defined epistemic equilibrium models (EEMs) first, by means of a truth-minimising criterion and then autoepistemic equilibrium models (AEEMs) by maximising ignorance, in the spirit of existing nonmonotonic epistemic logics. For both semantics we have provided strong equivalence characterisations. Our main contribution is that AEEMs provide an interesting logical semantics not only for epistemic specifications, but also for programs with arbitrary nestings of  $K$  and  $\rightarrow$ .

We have compared our semantics with the existing semantics for epistemic specifications. Overall, our approach provides the intended results when the program has two parts: one that produces several possible answer sets such as  $p \text{ or } q \leftarrow$  and another part where the  $M$ 's and  $K$ 's determine which of these answer sets is preferable, such as  $p \leftarrow \text{ not } K p$ . This holds for Gelfond's well-known example about student eligibility, and it seems that such cases were Gelfond's original motivation. Our semantics basically behaves in the same way as Kahl et al.'s for simple examples as in Table 5 and Table 6, which include cyclic dependences like  $p \leftarrow M p$ . We have however seen that there are subtle differences for more complex examples, and the controversial discussions of some examples in the literature show that it is not always clear what the result should be. We believe that the principled design of EHT models, EEMs and AEEMs speaks in favour of our approach. First, our EHT models interpreting two non-dual modal operators  $K$  and  $\hat{K}$  are in the tradition of intuitionistic modal logics. Second, the minimisation condition for EEMs naturally extends that of Pearce's equilibrium models. Third, the two minimisation conditions for AEEMs aim to capture Gelfond's idea of quantifying over possible answer sets. Among these two conditions, the subset maximality condition is certainly natural, while alternative definitions of the preference orderings  $\leq_\varphi$  might be designed.

<sup>13</sup> Recall that [3] computes two world views  $\{\emptyset\}$  and  $\{\{p\}\}$  for both  $p \leftarrow K p$  and  $p \leftarrow M p$ . For the former rule,  $\{\{p\}\}$  is counterintuitive, and this result has been justified by all semantics proposals in the literature. As for the latter, having 2 world views is unintended; while most of the semantics proposals in the literature support  $\{\{p\}\}$ , there are also a few supporting the other world view  $\{\emptyset\}$ , but there is none except [3] supporting both of them.

<sup>14</sup> Remember that [10] produces a world view  $\{\emptyset\}$  for the program  $p \leftarrow p, \text{ not } p$ , which departs even from the standard semantics of ASP programs.

We believe that epistemic here-and-there logic EHT provides a suitable basis for further extensions of answer-set programs with modal concepts, which is a research avenue that has recently arisen some interest specifically for extensions by temporal [42–47] and dynamic [48] concepts. A first step could be the extension from single-agent epistemic logic to multi-agent epistemic logic.

We have recently proposed a unifying mechanism, capturing ASP, nonmonotonic KD45 (aka, autoepistemic logic), nonmonotonic SW5 (aka, reflexive autoepistemic logic) and nonmonotonic S4F [49–51] in a monotonic framework. As future work, we also would like to classify epistemic equilibrium logic under this approach so that we can compare all such nonmonotonic formalisms in a single monotonic setting.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

We wish to express our thanks to David Pearce, Michael Gelfond, Patrick Thor Kahl, Pedro Cabalar, Jorge Fandinno, Yan Zhang and Thomas Eiter for useful discussions and feedback. We also would like to thank the anonymous reviewers for their valued comments and corrections on the drafts of this work.

The work of Ezgi Iraz Su was supported in part by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) in project number 389792660 (TRR 248, Center for Perspicuous Systems).

## References

- [1] C. Baral, M. Gelfond, Logic programming and knowledge representation, *J. Log. Program.* 19 (1994) 73–148.
- [2] M. Gelfond, V. Lifschitz, Classical negation in logic programs and disjunctive databases, *New Gener. Comput.* 9 (3/4) (1991) 365–386.
- [3] M. Gelfond, Strong introspection, in: T.L. Dean, K.R. McKeown (Eds.), *Proceedings of the 9th National Conference on Artificial Intelligence*, vol. 1, Anaheim, CA, USA, 1991, AAAI Press/the MIT Press, 1991, pp. 386–391, <http://www.aaai.org/Library/AAAI/1991/aaai91-060.php>.
- [4] M. Gelfond, Logic programming and reasoning with incomplete information, *Ann. Math. Artif. Intell.* 12 (1–2) (1994) 89–116.
- [5] M. Gelfond, New semantics for epistemic specifications, in: J.P. Delgrande, W. Faber (Eds.), *Logic Programming and Nonmonotonic Reasoning – 11th International Conference, LPNMR 2011, Vancouver, Canada, May 16–19, 2011. Proceedings*, in: *Lecture Notes in Computer Science*, vol. 6645, Springer, 2011, pp. 260–265.
- [6] P.T. Kahl, Refining the semantics for epistemic logic programs, Ph.D. thesis, Texas Tech University, Department of Computer Science, Lubbock, TX, USA, May 2014.
- [7] P.T. Kahl, A.P. Leclerc, T.C. Son, A parallel memory-efficient epistemic logic program solver: harder, better, faster, *CoRR*, arXiv:1608.06910 [abs], 2016.
- [8] J. Chen, The generalized logic of only knowing (GOL) that covers the notion of epistemic specifications, *J. Log. Comput.* 7 (2) (1997) 159–174.
- [9] K. Wang, Y. Zhang, Nested epistemic logic programs, in: C. Baral, G. Greco, N. Leone, G. Terracina (Eds.), *Logic Programming and Nonmonotonic Reasoning, 8th International Conference, LPNMR 2005, Diamante, Italy, September 5–8, 2005. Proceedings*, in: *Lecture Notes in Computer Science*, vol. 3662, Springer, 2005, pp. 279–290.
- [10] M. Truszczynski, Revisiting epistemic specifications, in: M. Balduccini, T.C. Son (Eds.), *Logic Programming, Knowledge Representation, and Nonmonotonic Reasoning – Essays Dedicated to Michael Gelfond on the Occasion of His 65th Birthday*, in: *Lecture Notes in Computer Science*, vol. 6565, Springer, 2011, pp. 315–333.
- [11] P. Kahl, R. Watson, E. Balai, M. Gelfond, Y. Zhang, The language of epistemic specifications (refined) including a prototype solver, *J. Log. Comput.* (2015), <https://doi.org/10.1093/logcom/exv065>.
- [12] Y. Shen, T. Eiter, Evaluating epistemic negation in answer set programming, *Artif. Intell.* 237 (2016) 115–135, <https://doi.org/10.1016/j.artint.2016.04.004>.
- [13] T.C. Son, T. Le, P.T. Kahl, A.P. Leclerc, On computing world views of epistemic logic programs, *Sierra* [53], pp. 1269–1275, <https://doi.org/10.24963/ijcai.2017/176>.
- [14] P.T. Kahl, A.P. Leclerc, Epistemic logic programs with world view constraints, in: A.D. Palù, P. Tarau, N. Saeedloei, P. Fodor (Eds.), *Technical Communications of the 34th International Conference on Logic Programming, ICLP 2018, July 14–17, 2018, Oxford, United Kingdom*, in: *OASICS*, vol. 64, Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2018, pp. 1:1–1:17, <https://doi.org/10.4230/OASICS.ICLP.2018.1>.
- [15] D. Pearce, A new logical characterisation of stable models and answer sets, in: J. Dix, L.M. Pereira, T.C. Przymusiński (Eds.), *Non-Monotonic Extensions of Logic Programming, NMELP '96, Bad Honnef, Germany, September 5–6, 1996. Selected Papers*, in: *Lecture Notes in Computer Science*, vol. 1216, Springer, 1996, pp. 57–70, <https://doi.org/10.1007/BFb0023801>.
- [16] D. Pearce, Equilibrium logic, *Ann. Math. Artif. Intell.* 47 (1–2) (2006) 3–41.
- [17] A. Heyting, Die formalen Regeln der intuitionistischen Logik, *Sitzber. Preuss. Akad. Wiss.* 42 (71) (1930) 158–169.
- [18] V. Lifschitz, D. Pearce, A. Valverde, Strongly equivalent logic programs, *ACM Trans. Comput. Log.* 2 (4) (2001) 526–541.
- [19] L. Fariñas del Cerro, A. Herzig, E.I. Su, Epistemic equilibrium logic, in: Q. Yang, M.J. Wooldridge (Eds.), *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25–31, 2015*, AAAI Press, 2015, pp. 2964–2970, <http://ijcai.org/Abstract/15/419>.
- [20] Y. Shen, T. Eiter, Evaluating epistemic negation in answer set programming (extended abstract), in: *Sierra* [53], pp. 5060–5064, <https://doi.org/10.24963/ijcai.2017/722>.
- [21] E.I. Su, Extensions of equilibrium logic by modal concepts. (Extensions de la logique d'équilibre par des concepts modaux), Ph.D. thesis, Institut de Recherche en Informatique de Toulouse, France, 2015, <https://tel.archives-ouvertes.fr/tel-01636791>.
- [22] Y. Zhang, Y. Zhang, Epistemic specifications and conformant planning, in: *The Workshops of the Thirty-First AAAI Conference on Artificial Intelligence, Saturday, February 4–9, 2017*, in: *WS-17 of AAAI Workshops*, vol. WS-17, AAAI Press, San Francisco, California, USA, 2017, <http://aaai.org/ocs/index.php/WS/AAAIW17/paper/view/15073>.
- [23] P. Cabalar, J. Fandinno, L. Fariñas del Cerro, Splitting epistemic logic programs, in: Balduccini et al. [52], pp. 120–133, [https://doi.org/10.1007/978-3-030-20528-7\\_10](https://doi.org/10.1007/978-3-030-20528-7_10).



- [24] L.C. Aiello, F. Massacci, An executable specification language for planning attacks to security protocols, in: *Proceedings of the 13th IEEE Computer Security Foundations Workshop, CSFW '00*, July 3-5, 2000, IEEE Computer Society, Cambridge, England, UK, 2000, pp. 88–102.
- [25] L.C. Aiello, F. Massacci, Verifying security protocols as planning in logic programming, *ACM Trans. Comput. Log.* 2 (4) (2001) 542–580, <https://doi.org/10.1145/383779.383785>.
- [26] L.C. Aiello, F. Massacci, Planning attacks to security protocols: case studies in logic programming, in: A.C. Kakas, F. Sadri (Eds.), *Computational Logic: Logic Programming and Beyond, Essays in Honour of Robert A. Kowalski, Part I*, in: *Lecture Notes in Computer Science*, vol. 2407, Springer, 2002, pp. 533–560, [https://doi.org/10.1007/3-540-45628-7\\_20](https://doi.org/10.1007/3-540-45628-7_20).
- [27] V. Lifschitz, L.R. Tang, H. Turner, Nested expressions in logic programs, *Ann. Math. Artif. Intell.* 25 (3–4) (1999) 369–389.
- [28] G. Fischer-Servi, On modal logic with an intuitionistic base, *Stud. Log.* 36 (4) (1976) 141–149.
- [29] L. Fariñas del Cerro, A.R. Raggio, Some results in intuitionistic modal logic, *Log. Anal.* 26 (102) (1983) 219–224.
- [30] A.K. Simpson, The proof theory and semantics of intuitionistic modal logic, Ph.D. thesis, School of Informatics, University of Edinburgh, College of Science and Engineering, November 1994.
- [31] G.M. Bierman, V. de Paiva, On an intuitionistic modal logic, *Stud. Log.* 65 (3) (2000) 383–416.
- [32] R.C. Moore, Possible-world semantics for autoepistemic logic, in: *Proceedings of the Non-Monotonic Reasoning Workshop, Mohonk Mountain House, New Paltz, NY 12561, USA, October 17-19, 1984*, American Association for Artificial Intelligence (AAAI), 1984, pp. 344–354.
- [33] H.J. Levesque, All I know: a study in autoepistemic logic, *Artif. Intell.* 42 (2–3) (1990) 263–309.
- [34] P. Cabalar, J. Fandinno, L. Fariñas del Cerro, D. Pearce, A. Valverde, On the properties of atom definability and well-supportedness in logic programming, in: E.C. Oliveira, J. Gama, Z.A. Vale, H.L. Cardoso (Eds.), *Progress in Artificial Intelligence - 18th EPIA Conference on Artificial Intelligence, EPIA 2017, Porto, Portugal, September 5-8, 2017*, Proceedings, in: *Lecture Notes in Computer Science*, vol. 10423, Springer, 2017, pp. 624–636, [https://doi.org/10.1007/978-3-319-65340-2\\_51](https://doi.org/10.1007/978-3-319-65340-2_51).
- [35] P. Blackburn, M. de Rijke, Y. Venema, *Modal Logic*, Cambridge Tracts in Theoretical Computer Science, Cambridge University Press, 2001.
- [36] P. Cabalar, J. Fandinno, L. Fariñas del Cerro, Founded world views with autoepistemic equilibrium logic, in: Balduccini et al. [52], pp. 134–147, [https://doi.org/10.1007/978-3-030-20528-7\\_11](https://doi.org/10.1007/978-3-030-20528-7_11).
- [37] J. Fandinno, Founded (auto)epistemic equilibrium logic satisfies epistemic splitting, *Theory Pract. Log. Program.* 19 (5–6) (2019) 671–687, <https://doi.org/10.1017/S1471068419000127>.
- [38] E.I. Su, Epistemic answer set programming, in: F. Calimeri, N. Leone, M. Manna (Eds.), *Logics in Artificial Intelligence - 16th European Conference, JELIA 2019, Rende, Italy, May 7-11, 2019*, Proceedings, in: *Lecture Notes in Computer Science*, vol. 11468, Springer, 2019, pp. 608–626.
- [39] M. Morak, Epistemic logic programs: a different world view, in: B. Bogaerts, E. Erdem, P. Fodor, A. Formisano, G. Ianni, D. Inlezan, G. Vidal, A. Villanueva, M.D. Vos, F. Yang (Eds.), *Proceedings 35th International Conference on Logic Programming (Technical Communications), ICLP 2019 Technical Communications, Las Cruces, NM, USA, September 20–25, 2019*, vol. 306, 2019, pp. 52–64.
- [40] S. Costantini, About epistemic negation and world views in epistemic logic programs, *Theory Pract. Log. Program.* 19 (5–6) (2019) 790–807, <https://doi.org/10.1017/S147106841900019X>.
- [41] E.I. Su, Strongly equivalent epistemic answer set programs, in: M. Alviano, G. Greco, F. Scarcello (Eds.), *AI\*IA 2019 - Advances in Artificial Intelligence - XVIIIth International Conference of the Italian Association for Artificial Intelligence, Rende, Italy, November 19–22, 2019*, Proceedings, in: *Lecture Notes in Computer Science*, vol. 11946, Springer, 2019, pp. 101–115.
- [42] F. Aguado, P. Cabalar, M. Diéguez, G. Pérez, C. Vidal, Temporal equilibrium logic: a survey, *J. Appl. Non-Class. Log.* 23 (1–2) (2013) 2–24, <https://doi.org/10.1080/11663081.2013.798985>.
- [43] M. Diéguez, Temporal answer set programming, Ph.D. thesis, Univ. Corunna, Department of Computer Science, Univ. Corunna, Spain, Feb. 2015.
- [44] P. Balbiani, M. Diéguez, Temporal here and there, in: L. Michael, A.C. Kakas (Eds.), *Logics in Artificial Intelligence - 15th European Conference, JELIA 2016, Larnaca, Cyprus, November 9–11, 2016*, Proceedings, in: *Lecture Notes in Computer Science*, vol. 10021, 2016, pp. 81–96.
- [45] F. Aguado, P. Cabalar, M. Diéguez, G. Pérez, C. Vidal, Temporal equilibrium logic with past operators, *J. Appl. Non-Class. Log.* 27 (3–4) (2017) 161–177, <https://doi.org/10.1080/11663081.2018.1427987>.
- [46] P. Cabalar, R. Kaminski, T. Schaub, A. Schuhmann, Temporal answer set programming on finite traces, *Theory Pract. Log. Program.* 18 (3–4) (2018) 406–420, <https://doi.org/10.1017/S1471068418000297>.
- [47] A. Bosser, P. Cabalar, M. Diéguez, T. Schaub, Introducing temporal stable models for linear dynamic logic, in: M. Thielscher, F. Toni, F. Wolter (Eds.), *Principles of Knowledge Representation and Reasoning: Proceedings of the 16th International Conference, KR 2018, Tempe, Arizona, 30 October–2 November 2018*, AAAI Press, 2018, pp. 12–21, <https://aaai.org/ocs/index.php/KR/KR18/paper/view/18047>.
- [48] L. Fariñas del Cerro, A. Herzig, E.I. Su, Combining equilibrium logic and dynamic logic, in: P. Cabalar, T.C. Son (Eds.), *Logic Programming and Nonmonotonic Reasoning, 12th International Conference, LPNMR 2013, Corunna, Spain, September 15–19, 2013*, Proceedings, in: *Lecture Notes in Computer Science*, vol. 8148, Springer, 2013, pp. 304–316.
- [49] L. Fariñas del Cerro, A. Herzig, E.I. Su, Capturing equilibrium models in modal logic, *J. Appl. Log.* 12 (2) (2014) 192–207, <https://doi.org/10.1016/j.jal.2013.12.003>.
- [50] E.I. Su, A monotonic view on reflexive autoepistemic reasoning, in: M. Balduccini, T. Janhunen (Eds.), *Logic Programming and Nonmonotonic Reasoning - 14th International Conference, LPNMR 2017, Espoo, Finland, July 3–6, 2017*, Proceedings, in: *Lecture Notes in Computer Science*, vol. 10377, Springer, 2017, pp. 85–100.
- [51] E.I. Su, An S4F-related monotonic modal logic, in: D.D. Monica, A. Murano, S. Rubin, L. Sauro (Eds.), *Joint Proceedings of the 18th Italian Conference on Theoretical Computer Science and the 32nd Italian Conference on Computational Logic Co-located with the 2017 IEEE International Workshop on Measurements and Networking, 2017 IEEE M&N, Naples, Italy, September 26-28, 2017*, in: *CEUR Workshop Proceedings*, vol. 1949, 2017, pp. 346–360, <http://ceur-ws.org/Vol-1949/CILCpaper11.pdf>, CEUR-WS.org.
- [52] M. Balduccini, Y. Lierler, S. Woltran (Eds.), *Logic Programming and Nonmonotonic Reasoning - 15th International Conference, LPNMR 2019, Philadelphia, PA, USA, June 3–7, 2019*, Proceedings, *Lecture Notes in Computer Science*, vol. 11481, Springer, 2019.
- [53] C. Sierra (Ed.), *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017*, 2017, ijcai.org, <http://www.ijcai.org/Proceedings/2017/>.