The modal logic of equilibrium models

Luis Fariñas del Cerro and Andreas Herzig

University of Toulouse
IRIT, CNRS
http://www.irit.fr

Abstract. Here-and-there models and equilibrium models were investigated as a semantical framework for answer set programming by Pearce, Cabalar, Lifschitz, Ferraris and others. The semantics of equilibrium logic is indirect in that the notion of satisfiability is defined in terms of satisfiability in the logic of here-and-there. We here give a direct semantics of equilibrium logic, stated in terms of a modal language into which the language of equilibrium logic can be embedded.

Keywords: equilibrium logic, here-and-there models, bimodal logic, answer-set programming

1 Introduction

A here-and-there model (HT model) is made up of two sets of propositional variables H ('here') and T ('there') such that $H \subseteq T$. The logical language to talk about such models has connectives $\bot$, $\wedge$, $\vee$, and $\Rightarrow$. The latter is interpreted in a non-classical way and is therefore different from the material implication $\rightarrow$:

$$H, T \models \varphi \Rightarrow \psi \quad \text{iff} \quad H, T \models \varphi \rightarrow \psi \text{ and } T, T \models \varphi \rightarrow \psi$$

where $\rightarrow$ is interpreted just as in classical propositional logic. Such models were studied since Gödel in order to give semantics to an implication with strength between intuitionistic and material implication [7]. They were later investigated by Pearce, Cabalar, Lifschitz, Ferraris and others as the basis of equilibrium logic, which is a semantical framework for answer set programming [10, 9, 11, 2, 3, 6, 8]; we refer to the equilibrium logic website[1] for an overview.

Equilibrium models of a formula φ are defined in an indirect way that is based on HT models: an equilibrium model of φ is a set of propositional variables T such that

1. $T, T \models \varphi$, and
2. there is no HT model (H, T) such that H is strictly *weaker* than T and $H, T \models \varphi$,

where 'weaker' means that H is included in T. For example, $T = \emptyset$ is an equilibrium model of $p \Rightarrow \bot$ because (1) for the HT model $(\emptyset, \emptyset)$ we have $\emptyset, \emptyset \models p \Rightarrow \bot$ and because (2) there is no set H that is strictly included in the empty set.

We here give a direct semantics of equilibrium logic in terms of a modal language having two unary modal operators [T] and [S]. Roughly speaking, [T] allows to talk

[1] http://www.equilibriumlogic.net

about the there-world: a valuation that is at least as strong as the actual valuation; and [S] allows to talk about all here-worlds that are possible if we take the actual world as a there-world: it quantifies over all valuations that are weaker than the actual world. This language can clearly be interpreted on HT models; however, we here give a semantics in terms of Kripke models. We call our logic **MEM**: the Modal logic of Equilibrium Models.

We relate the language of equilibrium logic to our bimodal language by means of a translation *tr*. The main clause of the translation is:

$$tr(\varphi \Rightarrow \psi) = (tr(\varphi) \rightarrow tr(\psi)) \wedge [\mathrm{T}](tr(\varphi) \rightarrow tr(\psi))$$

A first attempt to relate equilibrium logic to modal logic in the style of the present approach was presented in [5] in terms of modal operators of contingency. The present paper improves over it by providing a complete axiomatisation of the modal logic of equilibrium models in terms of a bimodal language.

This paper is organised as follows. In Section 2 we introduce our modal logic of equilibrium models **MEM** both semantically and axiomatically. In Section 3 we recall the logic of here-and-there and equilibrium logic. In Section 4 we define the translation *tr* from the language of the logic of here-and-there to the language of **MEM**; we prove that for any formula φ, φ is HT valid if and only if $tr(\varphi)$ is **MEM** valid. This paves the way to the proof that φ is a consequence of χ in equilibrium logic if and only if the modal formula

$$[\mathrm{T}]\,(tr(\chi) \wedge [\mathrm{S}]\neg tr(\chi)) \rightarrow [\mathrm{T}]tr(\varphi)$$

is valid in **MEM**. Section 5 concludes.

2 The modal logic of equilibrium models: MEM

We now introduce the modal logic of equilibrium models **MEM** in the classical way: we start by defining its bimodal language and its semantics and then axiomatise its validities.

2.1 Language

Throughout the paper we suppose given a countably infinite set of propositional variables $\mathbb{P}$. The elements of $\mathbb{P}$ are noted p, q, etc. Our language $\mathcal{L}_{[\mathrm{T}],[\mathrm{S}]}$ is bimodal: it has two modal operators [T] and [S]. Precisely, $\mathcal{L}_{[\mathrm{T}],[\mathrm{S}]}$ is defined by the following grammar:

$$\varphi ::= p \mid \bot \mid \varphi \rightarrow \varphi \mid [\mathrm{T}]\varphi \mid [\mathrm{S}]\varphi$$

where p ranges over $\mathbb{P}$. $[\mathrm{T}]\varphi$ may be read "φ holds at the there world" and $[\mathrm{S}]\varphi$ may be read "φ holds at every (strictly) weaker world".

The set of propositional variables occurring in formula φ is noted $\mathbb{P}_\varphi$.

$\mathcal{L}_{[\mathrm{T}]}$ is the sublanguage of $\mathcal{L}_{[\mathrm{T}],[\mathrm{S}]}$ formulas without [S], i.e., $\mathcal{L}_{[\mathrm{T}]}$ formulas are built from [T] and the Boolean connectives only.

We employ the standard abbreviations of the Boolean connectives: $\top \overset{\text{def}}{=} \bot \rightarrow \bot$, $\neg\varphi \overset{\text{def}}{=} \varphi \rightarrow \bot$, $\varphi \vee \psi \overset{\text{def}}{=} \neg\varphi \rightarrow \psi$, and $\varphi \wedge \psi \overset{\text{def}}{=} \neg(\varphi \rightarrow \neg\psi)$. Moreover, $\langle\mathrm{T}\rangle\varphi$ and $\langle\mathrm{S}\rangle\varphi$ respectively abbreviate $\neg[\mathrm{T}]\neg\varphi$ and $\neg[\mathrm{S}]\neg\varphi$.

2.2 Kripke models

We interpret the formulas of our language $\mathcal{L}_{[T],[S]}$ in a class of Kripke models that has to satisfy particular constraints. We then give an axiomatisation of the validities of that class of models and prove its completeness.

Consider the class of Kripke models $M = \langle W, \mathcal{T}, \mathcal{S}, V \rangle$ such that

- W is a non-empty set of possible worlds;
- V is a valuation on W mapping possible worlds $w \in W$ to sets of propositional variables $V_w \subseteq \mathbb{P}$;
- $\mathcal{T}, \mathcal{S} \subseteq W \times W$ are relations on W such that:

(d)	for every w there is a $v \in W$ such that $w\mathcal{T}v$;
(alt)	for every w, if $w\mathcal{T}v$ and $w\mathcal{T}v'$ then $v = v'$;
(heredity)	for every w, u, if $w\mathcal{S}u$ then $V_u \subseteq V_w$;
(negatable)	for every w, for every finite $P, Q \subseteq V_w$ such that P is nonempty and $P \cap Q = \emptyset$, there is u such that: $w\mathcal{S}u$, $V_u \cap P = \emptyset$ and $Q \subseteq V_u$;
(mtrans)	for every w, u, v, if $w\mathcal{S}u$ and $u\mathcal{T}v$ then $w\mathcal{T}v$;
(wconv)	for every w, v, if $w\mathcal{T}v$ then $w = v$ or $v\mathcal{S}w$.

The first two constraints are about the relation $\mathcal{T}$, the next two are about the relation $\mathcal{S}$, and the last two are about both. Constraints (d) and (alt) say that at any world w there is exactly one possible world that is accessible via $\mathcal{T}$. The (heredity) constraint is just as the heredity constraint of intuitionistic logic, except that the intuitionistic relation is the inverse of $\mathcal{S}$. In the finite case, the (negatable) and the (heredity) constraints together say basically that for every w, the set of worlds that are accessible from w via the relation $\mathcal{S}$ contains all those worlds u whose valuations V_u are strictly included in V_w. The mixed transitivity constraint (mtrans) together with (d) and (alt) entails that in $\mathcal{S}$ connected parts of the graph M there is a unique there-world. The weak conversion constraint (wconv) says that $\mathcal{T}$ is contained in $\mathcal{S}^{-1} \cup id_W$, where id_W is the diagonal of W.

Let us denote by $\mathcal{T}(w)$ the unique world that is accessible from w via $\mathcal{T}$. The function $\mathcal{T}$ is well-defined because of constraints (d) and (alt). Note that (wconv) can then be reformulated as: $\mathcal{T}(w) = w$ or $\mathcal{T}(w)\mathcal{S}w$, for every w.

Proposition 1. *The following properties hold for every Kripke model that satisfies the constraints above.*

1. *For every w, $\mathcal{T}(\mathcal{T}(w)) = \mathcal{T}(w)$, i.e., $\mathcal{T}$ is an idempotent function.*
2. *For every w, u, if $w\mathcal{S}u$ then $\mathcal{T}(w) = \mathcal{T}(u)$.*
3. *For every w such that V_w is finite, the set $\{V_u : w\mathcal{S}u\}$ equals either $\{V : V \subseteq V_w\}$, or $\{V : V \subset V_w\}$.*

The last property is due to the (heredity) and the (negatable) constraints and says that for finite V_w, the set of valuations associated to the worlds that are accessible from w via $\mathcal{S}$ is either the set of subsets of V_w or the set of strict subsets of V_w: it equals either 2^{V_w} or $2^{V_w} \setminus \{V_w\}$. This will be used in the proof of Proposition 9.

2.3 Truth conditions

The truth conditions for our bimodal logic are standard. The relation $\mathcal{T}$ interprets [T] and $\mathcal{S}$ interprets [S]:

$$
\begin{array}{lll}
M, w \models p & \text{iff} & p \in V_w; \\
M, w \not\models \bot; & & \\
M, w \models \varphi \to \psi & \text{iff} & M, w \not\models \varphi \text{ or } M, w \models \psi; \\
M, w \models [\mathrm{T}]\varphi & \text{iff} & M, \mathcal{T}(w) \models \varphi; \\
M, w \models [\mathrm{S}]\varphi & \text{iff} & M, u \models \varphi \text{ for every } u \text{ such that } w\mathcal{S}u.
\end{array}
$$

We say that φ has a Kripke model when $M, w \models \varphi$, for some model M and world w in M. We also say that φ is *satisfiable in Kripke models*. Moreover, φ is *valid in Kripke models* if and only if $M, w \models \varphi$ for every model M and possible world w of M.

The next proposition says that when checking satisfaction it is enough to only consider models with finite valuations.

Proposition 2. *Let φ be a $\mathcal{L}_{[\mathrm{T}],[\mathrm{S}]}$ formula. Let $M = \langle W, \mathcal{T}, \mathcal{S}, V\rangle$ be a Kripke model satisfying (d), (alt), (heredity), (negatable), (mtrans), and (wconv). Let the valuation V^φ be defined as follows:*

$$V^\varphi_w = V_w \cap \mathbb{P}_\varphi, \text{ for every } w \in W$$

Then $M^\varphi = \langle W, \mathcal{T}, \mathcal{S}, V^\varphi\rangle$ is a Kripke model satisfying (d), (alt), (heredity), (negatable), (mtrans), and (wconv), and $M, w \models \varphi$ if and only if $M^\varphi, w \models \varphi$.

Proof. That $M, w \models \varphi$ if and only if $M^\varphi, w \models \varphi$ can be shown by straightforward induction on the form of φ.

As to the constraints, those that are only about the accessibility relations are clearly preserved because we just modify the valuation. The model M^φ satisfies constraint (heredity): suppose $w\mathcal{S}u$; as M satisfies (heredity) we have $V_u \subseteq V_w$; hence $V^\varphi_u \subseteq V^\varphi_w$. Finally, the model M^φ satisfies (negatable): suppose $P, Q \subseteq V^\varphi_w = V_w \cap \mathbb{P}_\varphi$ are finite sets such that $P \neq \emptyset$; as M satisfies (negatable) there is u such that $w\mathcal{S}u$ and $V_u \cap P = \emptyset$ and $Q \subseteq V_u$. Clearly, for that u we also have $V^\varphi_u \cap P = \emptyset$; and for that very u we also have $Q \subseteq V^\varphi_u = V_u \cap \mathbb{P}_\varphi$. q.e.d.

We note that this property is different from the standard finite model property of modal logics which requires a finite set of possible worlds.

2.4 Axiomatics, decidability, and complexity

We now give an axiomatisation of the **MEM** validities.

First we define the fragment of *positive Boolean formulas* of $\mathcal{L}_{[\mathrm{T}],[\mathrm{S}]}$ by the following grammar:

$$\varphi^+ ::= p \mid \varphi^+ \wedge \varphi^+ \mid \varphi^+ \vee \varphi^+$$

Observe that every positive formula is falsifiable. (Note that $\top$ is not a positive Boolean formula.)

K([T])	the axioms and inference rules of modal logic **K** for [T]	
K([S])	the axioms and inference rules of modal logic **K** for [S]	
D([T])	$[\mathrm{T}]\varphi \rightarrow \langle\mathrm{T}\rangle\varphi$	
Alt([T])	$\langle\mathrm{T}\rangle\varphi \rightarrow [\mathrm{T}]\varphi$	
Heredity([S])	$\langle\mathrm{S}\rangle\varphi^+ \rightarrow \varphi^+$	for φ^+ a positive Boolean formula
Negatable([S])	$(\varphi^+ \wedge \psi) \rightarrow \langle\mathrm{S}\rangle(\neg\varphi^+ \wedge \psi)$	for φ^+ a pos. Boolean formula s.t. $\mathbb{P}_{\varphi^+} \cap \mathbb{P}_\psi = \emptyset$
MTrans([T], [S])	$[\mathrm{T}]\varphi \rightarrow [\mathrm{S}][\mathrm{T}]\varphi$	
WConv([T], [S])	$\varphi \rightarrow [\mathrm{T}](\varphi \vee \langle\mathrm{S}\rangle\varphi)$	

Table 1. Axiomatisation of **MEM**

Our axiom schemas and inference rules are listed in Table 1. The axiom schemas D([T]) and Alt([T]) are familiar from standard textbooks on modal logic. The schema Heredity([S]) captures the heredity constraint of intuitionistic logic. Note that it could be replaced by the axiom schema $\langle\mathrm{S}\rangle p \rightarrow p$, for p a propositional variable, or by $\neg\varphi^+ \rightarrow [\mathrm{S}]\neg\varphi^+$, for φ^+ a positive Boolean formula. The schema Negatable([S]) ensures that the modal operator [S] quantifies over all strict subsets of the actual valuation. The schema MTrans([T], [S]) is an axiom of mixed transitivity. The schema WConv([T], [S]) is a weak conversion axiom familiar from tense logics.

The notions of a proof and of *provability of a formula* are defined as usual in modal logic. For example $[\mathrm{S}]\bot \rightarrow \neg p$ can be proved from Negatable([S]) by **K**([S]), i.e., by standard modal principles. The proof of the transitivity axiom $[\mathrm{T}]\varphi \rightarrow [\mathrm{T}][\mathrm{T}]\varphi$ and its converse is a bit longer.

Proposition 3. *The schema* $[\mathrm{T}]\varphi \leftrightarrow [\mathrm{T}][\mathrm{T}]\varphi$ *is provable.*

Proof.

1. $[\mathrm{T}]\varphi \rightarrow [\mathrm{T}]([\mathrm{T}]\varphi \vee \langle\mathrm{S}\rangle[\mathrm{T}]\varphi)$ (axiom WConv([T], [S]))
2. $\langle\mathrm{S}\rangle[\mathrm{T}]\varphi \rightarrow \langle\mathrm{S}\rangle\langle\mathrm{T}\rangle\varphi$ (axiom D([T]) and **K**([S]))
3. $\langle\mathrm{S}\rangle\langle\mathrm{T}\rangle\varphi \rightarrow \langle\mathrm{T}\rangle\varphi$ (axiom MTrans([S], [T]))
4. $\langle\mathrm{T}\rangle\varphi \rightarrow [\mathrm{T}]\varphi$ (axiom Alt([T]))
5. $\langle\mathrm{S}\rangle[\mathrm{T}]\varphi \rightarrow [\mathrm{T}]\varphi$ (from 2, 3, 4)
6. $[\mathrm{T}]\varphi \rightarrow [\mathrm{T}]([\mathrm{T}]\varphi \vee [\mathrm{T}]\varphi)$ (from 1 and 5)
7. $[\mathrm{T}]\varphi \rightarrow [\mathrm{T}][\mathrm{T}]\varphi$ (from 6)
8. $[\mathrm{T}][\mathrm{T}]\varphi \rightarrow \langle\mathrm{T}\rangle\langle\mathrm{T}\rangle\varphi$ (axiom D([T]) twice, and **K**([T]))
9. $\langle\mathrm{T}\rangle\varphi \rightarrow \langle\mathrm{T}\rangle\langle\mathrm{T}\rangle\varphi$ (from 4, 7, 8)
10. $[\mathrm{T}]\varphi \leftrightarrow [\mathrm{T}][\mathrm{T}]\varphi$ (from 7, 9)

q.e.d.

The next schema is also going to be useful.

Proposition 4. *The following formula schema is provable:*

$$\textit{Negatable}'([\mathrm{S}]) \quad \big((\textstyle\bigwedge_{p\in P} p) \wedge (\bigwedge_{q\in Q} q)\big) \to \langle \mathrm{S}\rangle \big((\bigwedge_{p\in P} \neg p) \wedge (\bigwedge_{q\in Q} q)\big)$$
$$\textit{for } P, Q \subseteq \mathbb{P} \textit{ finite, } P \textit{ nonempty, and } P \cap Q = \emptyset$$

Proof. Negatable′([S]) can be proved from Negatable([S]) as follows. Suppose $P, Q \subseteq \mathbb{P}$ finite, P nonempty, and $P \cap Q = \emptyset$. The implication

$$((\bigwedge_{p\in P} p) \wedge (\bigwedge_{q\in Q} q)) \to ((\bigvee_{p\in P} p) \wedge (\bigwedge_{q\in Q} q))$$

is valid in classical propositional logic. Then Negatable′([S]) follows with the axiom schema Negatable([S]). q.e.d.

Our axiomatisation is sound and complete w.r.t. the set of formulas that are **MEM** valid.

Theorem 1. *Let φ be a $\mathcal{L}_{[\mathrm{T}],[\mathrm{S}]}$ formula. φ is valid in Kripke models of* **MEM** *if and only if φ is provable from the axioms and inference rules of* **MEM**.

Proof.

Soundness is proved as usual. We just consider the case of axiom Negatable([S]). Let φ^+ be a positive Boolean formula such that $\mathbb{P}_{\varphi^+} \cap \mathbb{P}_\psi = \emptyset$. Suppose $M, w \models \varphi^+ \wedge \psi$. Put φ^+ in conjunctive normal form, and let $\kappa = (\bigvee P)$ be some clause of that CNF, for some $P \subseteq \mathbb{P}_{\varphi^+} \neq \emptyset$. (Observe that $P \neq \emptyset$ by the definition of positive formulas.) Let $P_w = P \cap V_w$. We have $P_w \neq \emptyset$ because $M, w \models \kappa$. Let $Q = V_w \cap \mathbb{P}_\psi$. As M satisfies the constraint (negatable), there is a $u \in W$ such that $w\mathcal{S}u$, $V_u \cap P_w = \emptyset$ and $Q \subseteq V_u$. Hence $M, u \not\models \kappa$, and therefore $M, u \not\models \varphi^+$. As $\mathbb{P}_{\varphi^+} \cap \mathbb{P}_\psi = \emptyset$ and as V_u differs from V_w only by variables from $\mathbb{P}_{\varphi^+}$ we also have $M, u \models \psi$. Hence $M, u \models \neg\varphi^+ \wedge \psi$, and therefore $M, w \models \langle \mathrm{S}\rangle(\neg\varphi^+ \wedge \psi)$.

To prove completeness w.r.t. Kripke models of **MEM** we use canonical models [1, 4]. Consider the set W of maximal consistent sets of **MEM**. Define the accessibility relations $\mathcal{T}$ and $\mathcal{S}$ on W by:

$$u\mathcal{T}w \text{ iff } \{\varphi : [\mathrm{T}]\varphi \in u\} \subseteq w$$
$$u\mathcal{S}w \text{ iff } \{\varphi : [\mathrm{S}]\varphi \in u\} \subseteq w$$

and define a valuation V such that $V_w = w \cap \mathbb{P}$ for every $w \in W$. Let us prove that the canonical model is a legal Kripke model of **MEM**.

- Axioms D([T]) and Alt([T]) ensure that $\mathcal{T}$ is a total function, i.e., the canonical model satisfies constraints (d) and (alt).
- Axiom Heredity([S]) ensures that the canonical model satisfies the heredity constraint, viz. that $w\mathcal{S}u$ implies $V_u \subseteq V_w$. Indeed, suppose $w\mathcal{S}u$ and $p \in V_u = u$. As w contains $\langle \mathrm{S}\rangle p \to p$ and is maximal consistent we have $p \in w = V_w$.

– Axiom Negatable([S]) guarantees the (negatable) constraint. To see this take any $w \in W$ with $V_w \neq \emptyset$ and two finite sets of propositional variables $P, Q \subseteq w$ such that P is nonempty and $P \cap Q = \emptyset$. As w is a maximal consistent set it contains $(\bigwedge_{p \in P} p) \wedge (\bigwedge_{q \in Q} q)$. As by Proposition 4 w contains every instance of Negatable′([S]), it must also contain $\langle S \rangle((\bigwedge_{p \in P} \neg p) \wedge (\bigwedge_{q \in Q} q))$. Hence by definition of $\mathcal{S}$ there is some $u \in W$ such that $w\mathcal{S}u$ and u contains $(\bigwedge_{p \in P} \neg p) \wedge (\bigwedge_{q \in Q} q)$. Therefore $P \cap u = \emptyset$ and $Q \subseteq u$.
– The weak conversion axiom WConv([T], [S]) ensures constraint (wconv).
– The mixed transitivity axiom MTrans([T], [S]) ensures constraint (mtrans).

Hence the canonical model satisfies all constraints, and is therefore a legal Kripke model of **MEM**.

The proof of the truth lemma is as usual. q.e.d.

3 HT logic and equilibrium logic

In this section we recall HT logic and equilibrium logic.

3.1 The language $\mathcal{L}_{\Rightarrow}$

The language $\mathcal{L}_{\Rightarrow}$ is common to HT logic and equilibrium logic. It is defined by the following grammar:

$$\varphi ::= p \mid \bot \mid \varphi \wedge \varphi \mid \varphi \vee \varphi \mid \varphi \Rightarrow \varphi$$

where p ranges over $\mathbb{P}$. The other Boolean connectives are defined as abbreviations in the same way as for our bimodal language: negation $\neg\varphi$ is defined as $\varphi \Rightarrow \bot$, and $\top$ is defined as $\bot \Rightarrow \bot$.

3.2 Here-and-there logic

A *HT model* is a couple (H, T) such that $H \subseteq T \subseteq \mathbb{P}$. The set T is called 'there' and H is called 'here'.

Let (H, T) be a HT model. The truth conditions for $\mathcal{L}_{\Rightarrow}$ formulas are as follows:[2]

$H, T \models p$	iff	$p \in H$
$H, T \not\models \bot$		
$H, T \models \varphi \wedge \psi$	iff	$H, T \models \varphi$ and $H, T \models \psi$
$H, T \models \varphi \vee \psi$	iff	$H, T \models \varphi$ or $H, T \models \psi$
$H, T \models \varphi \Rightarrow \psi$	iff	$H, T \models \varphi \rightarrow \psi$ and $T, T \models \varphi \rightarrow \psi$

When $H, T \models \varphi$ then we say that (H, T) is a HT model of φ. A formula φ is *HT valid* if and only if every HT model is also a HT model of φ.

Proposition 5. *Let φ be a $\mathcal{L}_{\Rightarrow}$ formula and let q be a propositional variable such that $q \notin \mathbb{P}_{\varphi}$. Then $H, T \models \varphi$ iff $H, T \cup \{q\} \models \varphi$ iff $H \cup \{q\}, T \cup \{q\} \models \varphi$.*

[2] In the last clause we use material implication '$\rightarrow$' as a shorthand in order to give a concise formulation. To spell this out, its truth condition is: $H, T \models \varphi \rightarrow \psi$ iff $H, T \not\models \varphi$ or $H, T \models \psi$.

3.3 Equilibrium logic

An *equilibrium model* of a $\mathcal{L}_{\Rightarrow}$ formula φ is a set of propositional variables $T \subseteq \mathbb{P}$ such that

1. (T, T) is a HT model of φ;
2. no (H, T) with $H \subset T$ is a HT model of φ.

Here are three examples. First, the empty set is the only equilibrium model of both $\top$ and $\neg p$: for any $q \in \mathbb{P}$, $\{q\}$ is neither an equilibrium model of $\top$ nor of $\neg p$. Second, the set $\{p\}$ is *not* an equilibrium model of $\neg p \Rightarrow q$ because $\emptyset, \{p\} \models \neg p \Rightarrow q$. Third, $\{q\}$ is an equilibrium model of $\neg p \Rightarrow q$ because $\{q\}, \{q\} \models \neg p \Rightarrow q$ and $\emptyset, \{q\} \not\models \neg p \Rightarrow q$.

Let φ and χ be $\mathcal{L}_{\Rightarrow}$ formulas. φ is a *consequence of* χ *in equilibrium models*, written $\chi \models_{HT^*} \varphi$, if and only if for every equilibrium model T of χ, (T, T) is an HT model of φ. For example we have $\top \models_{HT^*} \neg p$ and $\neg p \Rightarrow q \models_{HT^*} q$.

4 From HT logic and equilibrium logic to modal logic

In this section we are going to translate HT logic and equilibrium logic into our logic **MEM**.

4.1 Translating $\mathcal{L}_{\Rightarrow}$ to $\mathcal{L}_{[\mathrm{T}]}$

To start we translate the language $\mathcal{L}_{\Rightarrow}$ of both HT logic and equilibrium logic into the language $\mathcal{L}_{[\mathrm{T}]}$ of **MEM**. We recursively define the mapping *tr* as follows:

$$\begin{array}{ll} tr(p) & = p \quad \text{for } p \in \mathbb{P} \\ tr(\bot) & = \bot \\ tr(\varphi \wedge \psi) & = tr(\varphi) \wedge tr(\psi) \\ tr(\varphi \vee \psi) & = tr(\varphi) \vee tr(\psi) \\ tr(\varphi \Rightarrow \psi) & = (tr(\varphi) \rightarrow tr(\psi)) \wedge [\mathrm{T}](tr(\varphi) \rightarrow tr(\psi)) \end{array}$$

This translation combines the Gödel translation from intuitionistic logic to modal logic **S4** with Boolos's splitting translation from modal logic **S4** to modal logic **K4**. The main clause of the former is $tr(\varphi \Rightarrow \psi) = \Box(tr(\varphi) \rightarrow tr(\psi))$, for some **S4** operator $\Box$. The main clause of the latter is $tr(\Box\varphi) = tr(\varphi) \wedge [\mathrm{T}]tr(\varphi)$, where $[\mathrm{T}]$ is a **K4** operator (the operator of our bimodal logic).

Here are some examples.

$$tr(\top) = tr(\bot \Rightarrow \bot) = (\bot \rightarrow \bot) \wedge [\mathrm{T}](\bot \rightarrow \bot).$$

The latter is equivalent to $\top$ in any normal modal logic.

$$tr(\neg p) = tr(p \Rightarrow \bot) = (p \rightarrow \bot) \wedge [\mathrm{T}](p \rightarrow \bot).$$

This is equivalent to $\neg p \wedge [\mathrm{T}]\neg p$ in any normal modal logic.

$$tr(p \vee \neg p) = tr(p) \vee tr(p \Rightarrow \bot) = p \vee ((p \rightarrow \bot) \wedge [\mathrm{T}](p \rightarrow \bot)).$$

This is equivalent to $p \vee [\mathrm{T}]\neg p$ in any normal modal logic.

Observe that translated formulas may be exponentially longer than the original formulas.

Our translation will be used to relate both HT logic and equilibrium logic to **MEM**.

4.2 From HT logic to MEM

On HT models, the fragment $\mathcal{L}_{[T]}$ of the language $\mathcal{L}_{[T],[S]}$ is at least as expressive as $\mathcal{L}_{\Rightarrow}$, modulo the translation *tr*.

Proposition 6. *Let T be a set of propositional variables and let $M_T = \langle W, \mathcal{T}, \mathcal{S}, V\rangle$ be a quadruple such that:*

$$\begin{aligned} W &= 2^T; \\ V_h &= h, \quad \textit{for every } h \in W; \\ \mathcal{T} &= W \times \{T\}; \\ \mathcal{S} &= \supset. \end{aligned}$$

Then M_T is a **MEM** *model, and $H, T \models \varphi$ if and only if $M_T, H \models tr(\varphi)$, for every $H \subseteq T$ and for every $\mathcal{L}_{\Rightarrow}$ formula φ.*

So in the last line $\mathcal{S}$ is defined to be the strict superset relation on 2^T. For example for the HT model $(\emptyset, \emptyset)$ we obtain $M_\emptyset = \langle W, \mathcal{T}, \mathcal{S}, V\rangle$ with $W = \{\emptyset\}$, $\mathcal{T} = \{\langle\emptyset, \emptyset\rangle\}$, and $\mathcal{S} = \emptyset$; and for the HT model $(\emptyset, \{p\})$ we obtain $M_{\{p\}} = \langle W, \mathcal{T}, \mathcal{S}, V\rangle$ with $W = \{\emptyset, \{p\}\}$, $\mathcal{T} = \{\langle\emptyset, \{p\}\rangle, \langle\{p\}, \{p\}\rangle\}$, and $\mathcal{S} = \{\{p\}, \emptyset\}\}$.

Proof. First, M_T is a legal **MEM** model: M_T satisfies constraints (d), (alt), (heredity), (negatable), (mtrans), and (wconv). Second, one can prove by a straightforward induction on the form of φ that $H, T \models \varphi$ iff $M_T, H \models tr(\varphi)$, for every $H \subseteq T$. q.e.d.

Proposition 7. *Let $M = \langle W, \mathcal{T}, \mathcal{S}, V\rangle$ be a* **MEM** *model. Then $M, w \models tr(\varphi)$ if and only if $V_w, V_{\mathcal{T}(w)} \models \varphi$, for every $w \in W$ and for every $\mathcal{L}_{\Rightarrow}$ formula φ.*

Proof. As expected the proof is by induction on the form of φ. The only non trivial case is that of the intuitionistic implication $\psi_1 \Rightarrow \psi_2$. We have:

$$\begin{aligned} M, w \models tr(\psi_1 \Rightarrow \psi_2) \ &\text{iff} \ M, w \models tr(\psi_1) \to tr(\psi_2) \text{ and } M, \mathcal{T}(w) \models tr(\psi_1) \to tr(\psi_2) \\ &\text{iff} \ V_w, V_{\mathcal{T}(w)} \models \psi_1 \to \psi_2 \text{ and } V_{\mathcal{T}(w)}, V_{\mathcal{T}(w)} \models \psi_1 \to \psi_2 \\ &\qquad \text{(by I.H. and by Prop. 1.1)} \\ &\text{iff} \ V_w, V_{\mathcal{T}(w)} \models \psi_1 \Rightarrow \psi_2 \end{aligned}$$

q.e.d.

Theorem 2. *Let φ be a $\mathcal{L}_{\Rightarrow}$ formula. Then φ is HT valid if and only if $tr(\varphi)$ is* **MEM** *valid.*

Proof. This follows from Proposition 6 and Proposition 7. q.e.d.

4.3 From equilibrium logic to MEM

The same construction as for HT logic allows us to turn equilibrium models into **MEM** models.

Proposition 8. *Let $T \subseteq \mathbb{P}$ and let $M_T = \langle W, \mathcal{T}, \mathcal{S}, V\rangle$ be a quadruple such that:*

$$\begin{aligned} W &= 2^T;\\ V_h &= h, \quad \text{for every } h \in W;\\ \mathcal{T} &= W \times \{T\};\\ \mathcal{S} &= \supset \text{ (the superset relation).}\end{aligned}$$

Then M_T is a **MEM** *model, and T is an equilibrium model of φ if and only if $M_T, \mathcal{T}(H) \models tr(\varphi) \wedge [\mathrm{S}]\neg tr(\varphi)$, for every $H \subseteq T$ and for every $\mathcal{L}_{\Rightarrow}$ formula φ.*

Proof. As we have already seen in Proposition 6, M_T is a legal **MEM** model; we in particular have $\mathcal{T}(H) = T$ (cf. Proposition 1). It remains to prove that T is an equilibrium model of φ iff for every $H \subseteq T$ we have $M_T, \mathcal{T}(H) \models tr(\varphi) \wedge [\mathrm{S}]\neg tr(\varphi)$, where φ is any $\mathcal{L}_{\Rightarrow}$ formula. We have:

T is an equilibrium model of φ
iff $T, T \models \varphi$ and $H, T \not\models \varphi$ for every $H \subset T$
iff $M_T, T \models tr(\varphi)$ and $M_T, H \not\models tr(\varphi)$ for every $H \subset T$ (by Proposition 6)
iff $M_T, T \models tr(\varphi)$ and $M_T, H \models \neg tr(\varphi)$ for every H such that $T\mathcal{S}H$
(because $T\mathcal{S}H$ iff $H \subset T$)
iff $M_T, T \models tr(\varphi)$ and $M_T, T \models [\mathrm{S}]\neg tr(\varphi)$
iff $M_T, \mathcal{T}(H) \models tr(\varphi) \wedge [\mathrm{S}]\neg tr(\varphi)$ for every $H \subseteq T$
(because $T = \mathcal{T}(H)$ for every $H \in W = 2^T$)

q.e.d.

Proposition 9. *Let $M = \langle W, \mathcal{T}, \mathcal{S}, V\rangle$ be a* **MEM** *model and let $w \in W$. Let φ be any $\mathcal{L}_{\Rightarrow}$ formula and let $q \in \mathbb{P} \setminus \mathbb{P}_\varphi$ be a propositional variable not occurring in φ. Define the set $T(w) \subseteq \mathbb{P}$ as:*

$$T(w) = \begin{cases} V_{\mathcal{T}(w)} & \text{if } V_u \subset V_{\mathcal{T}(w)} \text{ for every } u \text{ such that } \mathcal{T}(w)\mathcal{S}u\\ V_{\mathcal{T}(w)} \cup \{q\} & \text{if } V_u = V_{\mathcal{T}(w)} \text{ for some } u \text{ such that } \mathcal{T}(w)\mathcal{S}u\end{cases}$$

Then $M, \mathcal{T}(w) \models tr(\varphi) \wedge [\mathrm{S}]\neg tr(\varphi)$ if and only if $T(w)$ is an equilibrium model of φ.

Proof. By Proposition 2 we may suppose w.l.o.g. that V_w is finite for every $w \in W$. We have two cases.

First, when $V_u \subset V_{\mathcal{T}(w)}$ for every u such that $\mathcal{T}(w)\mathcal{S}u$, then by Item 3 of Proposition 1, the set of the valuations of $\mathcal{S}$ accessible worlds equals the set of strict subsets of

$V_{\mathcal{T}(w)}$. Therefore:

$$\begin{aligned}
& M, \mathcal{T}(w) \models tr(\varphi) \wedge [\mathrm{S}]\neg tr(\varphi) \\
\text{iff } & M, \mathcal{T}(w) \models tr(\varphi) \text{ and } M, u \not\models tr(\varphi) \text{ for every } u \text{ such that } \mathcal{T}(w)\mathcal{S}u \\
\text{iff } & V_{\mathcal{T}(w)}, V_{\mathcal{T}(\mathcal{T}(w))} \models \varphi \text{ and } V_u, V_{\mathcal{T}(u)} \not\models \varphi \text{ for every } u \text{ such that } \mathcal{T}(w)\mathcal{S}u \\
& \hfill \text{(by Proposition 7)} \\
\text{iff } & V_{\mathcal{T}(w)}, V_{\mathcal{T}(w)} \models \varphi \text{ and } V_u, V_{\mathcal{T}(w)} \not\models \varphi \text{ for every } u \text{ such that } \mathcal{T}(w)\mathcal{S}u \\
& \hfill \text{(by Proposition 1.1 and 1.2)} \\
\text{iff } & V_{\mathcal{T}(w)}, V_{\mathcal{T}(w)} \models \varphi \text{ and } H, V_{\mathcal{T}(w)} \not\models \varphi \text{ for every } H \subset V_{\mathcal{T}(w)} \\
& \hfill \text{(because } \{V_u : \mathcal{T}(w)\mathcal{S}u\} = 2^{\mathcal{T}(w)} \setminus \mathcal{T}(w), v.s.) \\
\text{iff } & T(w), T(w) \models \varphi \text{ and } H, T(w) \not\models \varphi \text{ for every } H \subset T(w)
\end{aligned}$$

Second, when $V_u = V_{\mathcal{T}(w)}$ for some u such that $w\mathcal{S}u$ then we have $T(w) = V_{\mathcal{T}(w)} \cup \{q\}$. Therefore:

$$\begin{aligned}
& M, \mathcal{T}(w) \models tr(\varphi) \wedge [\mathrm{S}]\neg tr(\varphi) \\
\text{iff } & M, \mathcal{T}(w) \models tr(\varphi) \text{ and } M, u \not\models tr(\varphi) \text{ for every } u \text{ such that } \mathcal{T}(w)\mathcal{S}u \\
\text{iff } & V_{\mathcal{T}(w)}, V_{\mathcal{T}(\mathcal{T}(w))} \models \varphi \text{ and } V_u, V_{\mathcal{T}(u)} \not\models \varphi \text{ for every } u \text{ such that } \mathcal{T}(w)\mathcal{S}u \\
& \hfill \text{(by Proposition 7)} \\
\text{iff } & V_{\mathcal{T}(w)}, V_{\mathcal{T}(w)} \models \varphi \text{ and } V_u, V_{\mathcal{T}(w)} \not\models \varphi \text{ for every } u \text{ such that } \mathcal{T}(w)\mathcal{S}u \\
& \hfill \text{(by Proposition 1.1 and 1.2)} \\
\text{iff } & V_{\mathcal{T}(w)}, V_{\mathcal{T}(w)} \models \varphi \text{ and } H, V_{\mathcal{T}(w)} \not\models \varphi \text{ for every } H \subseteq V_{\mathcal{T}(w)} \hfill \text{(v.s.)} \\
\text{iff } & V_{\mathcal{T}(w)} \cup \{q\}, V_{\mathcal{T}(w)} \cup \{q\} \models \varphi \text{ and } H, V_{\mathcal{T}(w)} \cup \{q\} \not\models \varphi \text{ for every } H \subset V_{\mathcal{T}(w)} \cup \{q\} \\
& \hfill \text{(by Proposition 5)} \\
\text{iff } & T(w), T(w) \models \varphi \text{ and } H, T(w) \not\models \varphi \text{ for every } H \subset T(w)
\end{aligned}$$

q.e.d.

For example consider the set $T = \emptyset$ and the formula $\varphi = \top$. We have seen above that $\emptyset$ is the only equilibrium model of $\top$. Let M_T be the Kripke model as constructed in propositions 6 and 8. Then $M_T, T \models [\mathrm{T}]\,(tr(\top) \wedge [\mathrm{S}]\neg tr(\top))$. This can be seen by simplifying the latter:

$$\begin{aligned}
[\mathrm{T}]\,(tr(\top) \wedge [\mathrm{S}]\neg tr(\top)) &\leftrightarrow [\mathrm{T}]\,(\top \wedge [\mathrm{S}]\neg\top) \\
&\leftrightarrow [\mathrm{T}][\mathrm{S}]\bot
\end{aligned}$$

We are now ready for the grand finale where we capture equilibrium logic in our bimodal logic.

Theorem 3. *Let φ and χ be $\mathcal{L}_{\Rightarrow}$ formulas. Then $\chi \models_{HT^*} \varphi$ if and only if*

$$[\mathrm{T}]\,(tr(\chi) \wedge [\mathrm{S}]\neg tr(\chi)) \rightarrow [\mathrm{T}]tr(\varphi)$$

is **MEM** *valid.*

Proof. This follows from Proposition 8 and Proposition 9. q.e.d.

Let us consider an example. We have seen that $\top \models_{HT^*} \neg p$, i.e., that $\neg p$ is a consequence of $\top$ in equilibrium models. We have seen in Section 4.1 that $tr(\top)$ is equivalent to $\top$ and that $tr(\neg p)$ is equivalent to $\neg p \wedge [\mathrm{T}]\neg p$. Theorem 3 tells us that the formula $\varphi = [\mathrm{T}](tr(\top) \wedge [\mathrm{S}]\neg tr(\top)) \to [\mathrm{T}](tr(\neg p))$ must be provable from the axioms and inference rules of **MEM**. This can be established by the following sequence of equivalent formulas:

1. $[\mathrm{T}](tr(\top) \wedge [\mathrm{S}]\neg tr(\top)) \to [\mathrm{T}](tr(\neg p))$
2. $[\mathrm{T}](\top \wedge [\mathrm{S}]\neg\top) \to [\mathrm{T}](\neg p \wedge [\mathrm{T}]\neg p)$ (v.s.)
3. $[\mathrm{T}][\mathrm{S}]\bot \to ([\mathrm{T}]\neg p \wedge [\mathrm{T}][\mathrm{T}]\neg p)$ (by **K**([T]))
4. $[\mathrm{T}][\mathrm{S}]\bot \to ([\mathrm{T}]\neg p \wedge [\mathrm{T}]\neg p)$ (by Proposition 3)
5. $[\mathrm{T}][\mathrm{S}]\bot \to [\mathrm{T}]\neg p$

The last line is provable in our logic: indeed, we have seen that $[\mathrm{S}]\bot \to \neg p$ can be proved from Negatable([S]) by standard modal principles. From this we can prove the last formula in our list by standard modal principles. Therefore the original formula φ is provable in our logic.

5 Conclusion

In this paper we have investigated the modal logic **MEM** that is behind equilibrium logic. We have shown that a logic with two modal operators [T] and [S] allows to capture the minimisation that is only expressed in the metalanguage in the standard definition of equilibrium models. We have shown that **MEM** satisfiability is decidable and that can be checked in polynomial space. We have also given a sound and complete axiomatisation.

It remains to give a lower bound for the complexity of **MEM**. It also remains to design a translation from the language of equilibrium logic to that of our bimodal logic that avoids exponential growth of the formula length. This can however be done in a quite straightforward way by integrating a modal operator $[\mathrm{T}]_*$ whose truth condition in HT models is:

$$H, T \models [\mathrm{T}]_*\varphi \text{ iff } H, T \models \varphi \text{ and } T, T \models \varphi$$

In terms of Kripke models $[\mathrm{T}]_*$ is interpreted by the reflexive closure of the accessibility relation $\mathcal{T}$ interpreting [T]. However, a drawback of the addition of a third modal operator is that the formalism gets more cumbersome.

Acknowledgements

This work was partially supported by the French-Spanish *Laboratoire Européen Associé (LEA) "French-Spanish Lab of Advanced Studies in Information Representation and Processing"*.

Thanks are due to David Pearce and Levan Uridia for their explanations about equilibrium logic. Thanks are also due to the FroCoS 2011 reviewers whose comments helped to improve the paper. Ezgi Iraz Su pointed out several flaws in the paper as it appeared in the proceedings of FroCoS 2011 that are corrected here.

References

1. Patrick Blackburn, Maarten de Rijke, and Yde Venema. *Modal Logic*. Cambridge Tracts in Theoretical Computer Science. University Press, 2001.
2. Pedro Cabalar and Paolo Ferraris. Propositional theories are strongly equivalent to logic programs. *Theory and Practice of Logic Programming (TPLP)*, 7(6):745–759, 2007.
3. Pedro Cabalar, David Pearce, and Agustín Valverde. Minimal logic programs. In Verónica Dahl and Ilkka Niemelä, editors, *Proc. ICLP*, volume 4670 of *LNCS*, pages 104–118. Springer Verlag, 2007.
4. Walter A. Carnielli, Claudio Pizzi, and Juliana Bueno-Soler. *Modalities and Multimodalities*. Logic, Epistemology, and the Unity of Science. Springer Verlag, 2009.
5. Luis Fariñas del Cerro and Andreas Herzig. Contingency-based equilibrium logic. In Jim Delgrande and Wolfgang Faber, editors, *Logic Programming and Nonmonotonic Reasoning (LPNMR)*, number 6645 in LNCS, pages 223–228, http://www.springerlink.com/, mai 2011. Springer-Verlag.
6. Paolo Ferraris, Joohyung Lee, and Vladimir Lifschitz. A new perspective on stable models. In Manuela M. Veloso, editor, *IJCAI*, pages 372–379, 2007.
7. A. Heyting. Die formalen Regeln der intuitionistischen Logik. *Sitzungsber. preuss. Akad. Wiss.*, 42-71:158–169, 1930.
8. Vladimir Lifschitz. Thirteen definitions of a stable model. In Andreas Blass, Nachum Dershowitz, and Wolfgang Reisig, editors, *Fields of Logic and Computation*, volume 6300 of *Lecture Notes in Computer Science*, pages 488–503. Springer Verlag, 2010.
9. Vladimir Lifschitz, David Pearce, and Agustín Valverde. Strongly equivalent logic programs. *ACM Transactions on Computational Logic*, 2(4):526–541, 2001.
10. David Pearce. A new logical characterisation of stable models and answer sets. In Jürgen Dix, Luís Moniz Pereira, and Teodor C. Przymusinski, editors, *NMELP*, volume 1216 of *Lecture Notes in Computer Science*, pages 57–70. Springer Verlag, 1996.
11. David Pearce, Inman P. de Guzmán, and Agustín Valverde. A tableau calculus for equilibrium entailment. In Roy Dyckhoff, editor, *TABLEAUX*, volume 1847 of *LNCS*, pages 352–367. Springer Verlag, 2000.