Dynamic Epistemic Logic

Jelle Gerbrandy

ILLC/Department of Philosophy
Nieuwe Doelenstraat 15
1012 CP, Amsterdam
tel.: +31-20-525 4551
e-mail: `gerbrand@illc.uva.nl`

1 Introduction

This paper is the result of combining two traditions in formal logic: epistemic logic and dynamic semantics.

Dynamic semantics is a branch of formal semantics that is concerned with *change*, and more in particular with change of information. The main idea in dynamic semantics is that the meaning of a syntactic unit—be it a sentence of natural language or a computer program—is best described as the change it brings about in the state of a human being or a computer. The motivation for, and applications of this 'paradigm-shift' can be found in areas such as semantics of programming languages (cf. Harel, 1984), default logic (Veltman, 1996), pragmatics of natural language (Stalnaker, 1972) and of man-computer interaction, theory of anaphora (Groenendijk and Stokhof, 1991) and presupposition theory (Beaver, 1995). Van Benthem (1996) provides a survey.

This paper is firmly rooted in this paradigm, but at the same time it is much influenced by another tradition: that of the analysis of epistemic logic in terms of multi-modal Kripke models.

This paper is the result of combining these two traditions. It contains a semantics and a deduction system for a multi-agent modal language extended with a repertoire of programs that describe information change. The language is designed in such a way that everything that is expressible in the object language can be known or learned by each of the agents. The possible use of this system is twofold: it might be used as a tool for reasoning agents in computer science and it might be used as a logic for formalizing certain parts of pragmatics and discourse theory.[1]

[1] As a first step in this direction, Gerbrandy and Groeneveld (to appear) show how a logic similar to the one introduced in this paper can be used to formalize the puzzles like the

The paper is organized as follows. The next section contains a short description of classical modal logic and introduces models based on non-well-founded sets as an alternative to Kripke semantics. In the section after that I introduce programs and their interpretation and I give a sound and complete axiomatization of the resulting logic in section 4. The last section is devoted to a comparison with update semantics of Veltman (1996).

Finally, I would like to mention the dissertations of Groeneveld (1995), Jaspars (1994) and de Rijke (1992) and the book by Fagin, Halpern, Moses and Vardi (1995) as precursors and sources of inspiration. The article by Willem Groeneveld and me (to appear) contains some ideas similar to those presented here.

2 Static Modal Semantics

The classical language of multi-modal logic is the following:

Definition 2.1 Let $\mathcal{A}$ be a set of agents and $\mathcal{P}$ a set of propositional variables. The language of classical modal logic is given by:

$$\Phi ::= p \mid \phi \wedge \psi \mid \neg\phi \mid \Box_a \phi$$

where $p \in \mathcal{P}$ and $a \in \mathcal{A}$.

One way of providing a semantics for this language is in terms of Kripke models. A *pointed Kripke model* is a quadruple $\langle W, \{R_a\}_{a \in \mathcal{A}}, V, w\rangle$, where W is a set of possible worlds, w is a distinguished element of W (the point of evaluation), R_a is a relation on W for each $a \in \mathcal{A}$, V is a valuation function that assigns a truth-value (either 0 or 1) to each pair of a world $v \in W$ and a propositional variable $p \in \mathcal{P}$.

Intuitively, given a Kripke model and a world w in it, the information of an agent a in w is represented by the set of worlds that are accessible from w via R_a; these worlds are the worlds compatible with a's information in w.

Kripke models have been studied extensively and they provide a very perspicuous semantics for the classical language of epistemic logic. Unfortunately, it turns out that Kripke-models are not very suitable structures for defining operations that correspond to intuitive notions of information change.[2] To avoid this problem, I use a different (but equivalent) representation.

Definition 2.2 Possibilities
Let $\mathcal{A}$, a set of agents, and $\mathcal{P}$, a set of propositional variables, be given. The class of possibilities is the largest class such that:

Conway paradox or the puzzle of the dirty children.

[2]Cf. Groeneveld (1995) for a discussion of the problems one encounters.

- A possibility w is a function that assigns to each propositional variable $p \in \mathcal{P}$ a truth value $w(p) \in \{0, 1\}$ and to each agent $a \in \mathcal{A}$ an information state $w(a)$.
- An information state σ is a set of possibilities.

A possibility w characterizes which propositions are true and which are false, and it characterizes the information that each of the agents has in the form of an information state σ, that consists of the set of possibilities the agent considers possible in w.[3]

This definition of possibilities should be read to range over the universe of non-well-founded sets in the sense of Aczel (1988).[4] The form of this definition, defining a set co-inductively as 'the largest class such that....', is a standard form of definition in non-well-founded set theory.

Truth of classical modal sentences in a possibility can be defined in a way analogous to the definition of truth for Kripke models.

Definition 2.3 Truth.
Let w be a possibility.

$$\begin{array}{rcl} w \models p & \text{iff} & w(p) = 1 \\ w \models \phi \wedge \psi & \text{iff} & w \models \phi \text{ and } w \models \psi \\ w \models \neg\phi & \text{iff} & w \not\models \phi \\ w \models \Box_a \phi & \text{iff} & \text{for all } v \in w(a) : v \models \phi \end{array}$$

It turns out that using possibilities instead of Kripke-models does not make an essential logical difference: there is a close relation between possibilities and pointed Kripke models.

Definition 2.4 Let $\mathcal{K} = (W, \{R_a\}_{a \in \mathcal{A}}, V, w)$ be a pointed Kripke model.

- A *decoration* of $\mathcal{K}$ is a function d that assigns to each world $v \in W$ a function with $\mathcal{P} \cup \mathcal{A}$ as its domain, such that $d(v)(p) = V(v, p)$ for each $p \in \mathcal{P}$, and $d(v)(a) = \{d(u) \mid vR_a u\}$ for each $a \in \mathcal{A}$.
- If $\mathcal{K} = (W, \{R_a\}_{a \in \mathcal{A}}, V, w)$ is a Kripke model, and d is a decoration of it, $d(w)$ is its *solution*, and $\mathcal{K}$ is a *picture* of $d(w)$.

A decoration of a Kripke model assigns to each possible world w in the model a possibility that assigns the same truth-values to the propositional variables as they get in the model at w, and that assigns to each agent a the set of possibilities that are assigned to worlds accessible from w by R_a.

The notions of solution and picture give us a correspondence between Kripke-models and possibilities:

---

[3]In Aczel (1988) as well as in Barwise and Moss (1996) similar models are defined.

[4]To be precise, the underlying set-theory is axiomatized by ZFC^- (the Zermelo-Fraenkel axioms without the axiom of foundation) plus Aczel's Anti-Foundation Axiom (AFA).

Proposition 2.5

- Each Kripke model has a unique solution, which is a possibility.
- Each possibility has a Kripke model as its picture.
- Two Kripke-models are pictures of the same possibility iff they are bisimilar.

Defining truth of a formula in a Kripke model in the standard way, it holds that:

Proposition 2.6 For each possibility w:

$$w \models \phi \text{ iff } \phi \text{ is true in each picture of } w$$

So a possibility and a picture of it are descriptively equivalent. This means that one can see possibilities as representatives of equivalence classes of Kripke models under bisimulation.

3 Programs

In this section we will define operations on possibilities that correspond to changes in the information states of the agents. The kind of information change we want to model is that of agents getting new information or learning that the information state of some other agent has changed in a certain way. I will introduce 'programs' in the object language that describe such changes. Changes in the 'real world' will not be modeled, and I will ignore other operations of information change such as belief contraction or belief revision.

The programming language is built up as follows. There are programs of the form $?\phi$ for each sentence ϕ. A program of the form $?\phi$ will be interpreted as a test that succeeds in a possibility when ϕ is true and fails otherwise. The language contains a program operator U_a for each agent a. A program of the form $U_a\pi$ corresponds to agent a learning that program π has been executed. Finally, the language contains two operators that combine programs to form a new program: sequencing and disjunction. A program of the form $\pi; \pi'$ is interpreted as: "first execute π, then π'." Disjunction is interpreted as choice: $\pi \cup \pi'$ corresponds to executing either π or π'.

To connect the programming language to the 'static part' of the language, we add a modal operator $[\pi]$ for each program π. Intuitively, a sentence $[\pi]\psi$ is true in a possibility just in case that after executing the program π in that possibility, ψ must be true. The set of programs is defined simultaneously with the set of sentences in a way that might be familiar from propositional dynamic logic (cf. for example Pratt, 1976 or Goldblatt, 1987).

Definition 3.1 Language.
Given a set of agents $\mathcal{A}$ and a set of propositional variables $\mathcal{P}$, the set of sentences of dynamic epistemic logic is the smallest set given by:

$$\Phi ::= p \mid \phi \wedge \psi \mid \neg\phi \mid \Box_a \phi \mid [\pi]\phi$$

where $a \in \mathcal{A}$, $p \in \mathcal{P}$, and π is any program. The set of programs is the smallest set given by:

$$\Pi ::= ?\phi \mid U_a \pi \mid \pi ; \pi' \mid \pi \cup \pi'$$

Programs are interpreted as relations over possibilities: a pair of possibilities (w, v) will be in the denotation of a program π (written as $w[\![\pi]\!]v$) just in case the execution of the program π in possibility w may result in v. I propose the following definition (in the definition, I use the abbreviation $w[a]v$ for the statement that w differs at most from v in the state it assigns to a):

Definition 3.2 Interpretation of programs.

$$\begin{array}{rll} w[\![?\phi]\!]v & \text{iff} & w \models \phi \text{ and } w = v \\ w[\![U_a\pi]\!]v & \text{iff} & w[a]v \text{ and } v(a) = \{v' \mid \exists w' \in w(a) : w'[\![\pi]\!]v'\} \\ w[\![\pi;\pi']\!]v & \text{iff} & \text{there is a } u \text{ such that } w[\![\pi]\!]u[\![\pi']\!]v \\ w[\![\pi \cup \pi']\!]v & \text{iff} & \text{either } w[\![\pi]\!]v \text{ or } w[\![\pi']\!]v \end{array}$$

Furthermore, the definition of truth is extended with the following clause:

$$w \models [\pi]\phi \quad \text{iff} \quad \text{for all } v \text{ if } w[\![\pi]\!]v \text{ then } v \models \phi$$

Programs of the form $U_a\pi$ are to be read as "a learns that π has been executed," or, alternatively, as "a updates her information state with π." This is modeled as follows. Executing a program of the form $U_a\pi$ in a possibility w results in a new possibility v in which only a's information state has changed. The information state of a in v contains all and only those possibilities that are the possible result of an execution of π in one of the possibilities that in a's information state in w. Note that a program of the form $U_a\pi$ is deterministic; in fact, $[\![U_a\pi]\!]$ is always a total function, which means that the update always exists, and the result is unique.

In the case that π is a test of the form $?\phi$, the result of executing $U_a?\phi$ is such that in a's new information state, all possibilities in which ϕ is not true are discarded: the new information state of a contains only possibilities in which ϕ is true. So, one might say that $U_a?\phi$ corresponds to a getting the information that ϕ is the case.

The programming language is constructed in such a way that each program can be executed by each of the agents. This has the effect that any change in the model that we can express as a program in the object language can be 'learned' by each of the agents. In particular, this means that each sentence can be 'learned' by each of the agents, because there is a test $?\phi$ in the programming language for each sentence ϕ.

I will give some examples. A program of the form $U_a U_b ?p$ denotes the action that a updates her state with the information that b has updated his information state with $?p$. This corresponds with a getting the information that b has gotten the information that p is the case.

We can also express that a learns whether p is the case, a situation, for example, of an agent a checking the value of a bit (p expressing that the value is 0, $\neg p$ expressing it is 1) or of a philosopher looking out of the window to check whether it rains. This corresponds to the program $(?p; U_a?p) \cup (?\neg p; U_a?\neg p)$: if p is the case, a learns that p, and if p is not true, a learns that $\neg p$.

Conscious Updates

The resulting logic and semantics is very similar to the system 'Multi-agent Eliminative K' from Groeneveld (1995, p. 157 ff.). It suffers from the same kind of problems, most notably the fact that introspection is not preserved over U_a-updates. The problem is the following: if an agent a updates with π, she will change all the possibilities in her information state in the way the program tells her to. But each possibility in her information state also contains a representation of her own information, and this representation does not necessarily change: the possibilities in her new state will assign to a an information state that does not correspond to the information she actually has.

An example might make the matter more clear. Let the class of introspective possibilities be the largest class of possibilities w such that it holds that for each agent a, $v \in w(a)$ implies that $w(a) = v(a)$ and that v is an introspective possibility. So, a possibility is introspective just in case the information state of an agent a only contains possibilities in which a is assigned the information state she is actually in. If a possibility w is introspective, all sentences of the form $\Box_a \phi \to \Box_a \Box_a \phi$, and all sentences of the form $\neg \Box_a \phi \to \Box_a \neg \Box_a \phi$ are true. Introspection is a property that is often associated with knowledge or with belief. To give a plausible account of 'learning', one would like introspection to be a property that is preserved over updates.

Unfortunately, this is not the case with U_a-updates. For take an introspective possibility w and suppose that $w(a)$ contains both possibilities where p is true, and where p is not true, i.e. it holds that $w \models \neg \Box_a p$, $w \models \neg \Box_a \neg p$, and hence that $w \models \Box_a \neg \Box_a p$.

Consider now the possibility that results from updating w with $U_a?p$, i.e. the unique possibility v such that $w[\![U_a?p]\!]v$. This is a possibility in which a has got the information that p: $v \models \Box_a p$. But because each possibility in $v(a)$ also occurred in $w(a)$, $\neg \Box_a p$ is true in each possibility in $v(a)$. So, it holds that $v \models \Box_a \neg \Box_a p$, while it also holds that $v \models \Box_a p$.

To solve this problem, one would like that an update of a's information state with π not only changes each of a's possibilities in accord with π, but also changes each of these updated possibilities to the effect that she has learned π. It turns out that it is not very hard to define a notion of update which reflects this. I will refer to such an update as a 'conscious update,' because it reflects the idea that if a updates with π, she is conscious of this fact.[5] I use the notation $U_a^*\pi$ for a conscious update of a's information state with π.

Definition 3.3 Conscious update.

$$w[\![U_a^*\pi]\!]v \text{ iff } w[a]v \text{ and } v(a) = \{v' \mid \exists w' \in w(a) : w'[\![\pi]\!][\![U_a^*\pi]\!]v'\}$$

This definition is circular as it stands. Nevertheless, it is not very hard to prove that for each program π there is a unique relation $[\![U_a^*\pi]\!]$ that conforms to the definition.[6] Also, $[\![U_a^*\pi]\!]$ is a total function for each π.

The definition says that consciously updating a's information state in a possibility w with π results in a possibility v that differs only from w in that all possibilities in $w(a)$ are first updated with π, and after that with $U_a^*\pi$. That the interpretation of $U_a^*\pi$ gives the effect of a conscious update is corroborated by the fact that it holds that if a's information in w is introspective, it is introspective after the update of w with $U_a^*?\phi$.

Group Updates

Common knowledge is a concept that occurs under different names (mutual knowledge, common ground) in the literature. The usual definition is that a sentence ϕ is common knowledge in a group $\mathcal{B}$ just in case each agent in the group knows that ϕ is the case, each agent knows that each of the other agents knows that ϕ, etcetera. As Barwise (1989) shows, a semantics based on non-well-founded sets is quite useful for modeling this concept.

Instead of concentrating on this static notion of mutuality, I will introduce the notion of a 'group update': an update with a program π in a group of agents that has the effect of changing the state of each agent in the group in the way described by π in such a way that each agent in the group is aware of the fact that each agent has executed π, each agent knows that each agent in the group knows that π is executed by each agent in the group, etc. In case π is a test of the form $?\phi$, a common update with $?\phi$ corresponds to the sentence ϕ becoming common knowledge within the group.

To express this in the object language, I add program operators of the form $U_{\mathcal{B}}^*$ for each subset $\mathcal{B}$ of $\mathcal{A}$ to the language. They are interpreted as follows:

Definition 3.4 Group update
For each π and $\mathcal{B} \subseteq \mathcal{A}$:

$$w[\![U_{\mathcal{B}}^*\pi]\!]v \quad \text{iff} \quad w[\mathcal{B}]v \text{ and } \forall a \in \mathcal{B} :$$

[5] The terminology is from Groeneveld (1995), who introduces a notion of conscious update in a single agent setting.

[6] A proof of the correctness of a similar definition can be found in Gerbrandy and Groeneveld (to appear).

$$v(a) = \{v' \mid \exists w' \in w(a) : w' [\![\pi]\!] [\![U^*_{\mathcal{B}}\pi]\!] v'\}$$

Updating a possibility with a program $U^*_{\mathcal{B}}\pi$ results in a possibility v that differs only from w in that for each $a \in \mathcal{B}$, all situations in $w(a)$ are first updated with π, and then with $U^*_{\mathcal{B}}\pi$.

Note that a group update in a group consisting of a single agent boils down to the notion of a conscious update defined above: updating with $U^*_{\{a\}}\pi$ is exactly the same thing as updating with $U^*_a\pi$.

4 Axiomatization

The following set of axioms and rules provides a sound and complete characterization of the set of sentences that are true in all possibilities. (For sake of presentation, I have left out the conscious single agent updates, since they are a special case of the group updates with a group consisting of a single agent. I have also left out axioms for the non-conscious updates introduced in definition 3.2. The axioms for U_a are just as those for $U^*_{\{a\}}$, except for axiom 7, which should be changed into: $\vdash [U_a\pi]\Box_a\psi \leftrightarrow \Box_a[\pi]\psi$.)

Axioms

1 $\vdash \phi$ if ϕ is valid in classical propositional logic.

2 $\vdash \Box_a(\phi \to \psi) \to (\Box_a\phi \to \Box_a\psi)$

3 $\vdash [\pi](\phi \to \psi) \to ([\pi]\phi \to [\pi]\psi)$.

4 $\vdash [?\phi]\psi \leftrightarrow (\phi \to \psi)$

5 $\vdash \neg[U^*_{\mathcal{B}}\pi]\psi \leftrightarrow [U^*_{\mathcal{B}}\pi]\neg\psi$

6 $\vdash [U^*_{\mathcal{B}}\pi]p \leftrightarrow p$

7 $\vdash [U^*_{\mathcal{B}}\pi]\Box_a\phi \leftrightarrow \Box_a[\pi][U^*_{\mathcal{B}}\pi]\phi$ if $a \in \mathcal{B}$

8 $\vdash [U^*_{\mathcal{B}}\pi]\Box_a\phi \leftrightarrow \Box_a\phi$ if $a \notin \mathcal{B}$

9 $\vdash [\pi;\pi']\phi \leftrightarrow [\pi][\pi']\phi$

10 $\vdash [\pi \cup \pi']\psi \leftrightarrow ([\pi]\psi \wedge [\pi']\psi)$

Rules

MP $\phi, \phi \to \psi \vdash \psi$

Nec$\Box$ If $\vdash \phi$ then $\vdash \Box_a\phi$

Nec$[\cdot]$ If $\vdash \phi$ then $\vdash [\pi]\phi$

$\Gamma \vdash \phi$ iff there is derivation of ϕ from the premises in Γ using the rules and axioms above.

In addition to the rules and axioms of classical modal logic, the deduction system consists of axioms describing the behavior of the program operators.

Axiom 3 and the rule Nec[$\cdot$] guarantee that the program operators behave as normal modal operators. Axiom 4 says that performing a test $?\phi$ boils down to checking whether ϕ is true. Axiom 5 reflects the fact that U^*_a-updates are total functions: an update with $U^*_a\pi$ always gives a unique result. This means that if it is not the case that a certain sentence is true after an update with a program of the form $U_{\mathcal{B}}\pi$, then it must be the case that the negation of that sentence is true in the updated possibility, and vice versa. Axiom 6 expresses that the update of an information state has no effect on the 'real' world; the same propositional atoms will be true or false before and after an update. Axiom 7 expresses that after a group update with π, an agent in the group knows that ψ just in case that agent already knew that after executing π, an update with $U^*_{\mathcal{B}}\pi$ could only result in a possibility in which ψ were true. Axiom 8 expresses that a group update has no effect on the information of agents outside of that group. The axioms 9 and 10 govern the behavior of sequencing and disjunction respectively.

Proposition 4.1 Soundness
If $\Gamma \vdash \phi$ then $\Gamma \models \phi$.

proof: By a standard induction. By way of illustration, I will show the correctness of axiom 5. In the proof, I make use of the fact that $[\![U^*_{\mathcal{B}}\pi]\!]$ is a total function for each π, and write $w[\![U^*_{\mathcal{B}}\pi]\!]$ for the unique v such that $w[\![U^*_{\mathcal{B}}\pi]\!]v$. The following equivalences hold, if $a \in \mathcal{B}$:

$$
\begin{array}{lll}
w \models [U^*_{\mathcal{B}}\pi]\Box_a\phi & \text{iff} & w[\![U^*_{\mathcal{B}}\pi]\!] \models \Box_a\phi \\
& \text{iff} & \forall v \in w[\![U^*_{\mathcal{B}}\pi]\!](a) : v \models \phi \\
& \text{iff} & \forall v : \text{ if } \exists w' \in w(a) : w'[\![\pi]\!][\![U^*_{\mathcal{B}}\pi]\!]v \text{ then } v \models \phi \\
& \text{iff} & \forall w' \in w(a)\ \forall v : \text{ if } w'[\![\pi]\!][\![U^*_{\mathcal{B}}\pi]\!]v \text{ then } v \models \phi \\
& \text{iff} & \forall w' \in w(a) : w' \models [\pi][U^*_{\mathcal{B}}\pi]\phi \\
& \text{iff} & w \models \Box_a[\pi][U^*_{\mathcal{B}}\pi]\phi
\end{array}
$$

Proposition 4.2 Completeness
If $\Gamma \models \phi$, then $\Gamma \vdash \phi$.

proof: The completeness proof is rather long. I give here the main structure; the details are delegated to the appendix.

The proof is a variation on the classical Henkin proof for completeness of modal logic. It is easy to show that each consistent set can be extended to a maximal consistent set (this will be referred to as 'Lindenbaum's Lemma'). We must show that for each consistent set of sentences there is a possibility in which these sentences are true. Completeness then follows by a standard argument.

Let, for each maximal consistent set Σ, w_Σ be that possibility such that $w_\Sigma(p) = 1$ iff $p \in \Sigma$, and for each agent a: $w_\Sigma(a) = \{w_\Gamma \mid \Gamma$ is maximal consistent and if $\Box_a\psi \in \Sigma$, then $\psi \in \Gamma\}$.[7] We prove the usual truth lemma, namely that for each sentence ϕ it holds that $\phi \in \Sigma$ iff $w_\Sigma \models \phi$.

The truth lemma is proven by an induction on the structure of ϕ, in which all cases are standard, except the case where ϕ is of the form $[\pi]\psi$. The proof for this case rests on the following idea. Just as membership in $w_\Sigma(a)$ depends on the formulae of the form $\Box_a\phi$ in Σ, the π-update of w_Σ is closely related to the formulae of the form $[\pi]\psi$ in Σ. This is reflected by the following relation between maximal consistent sets:

$$\Sigma R_\pi \Gamma \quad \text{iff} \quad \Gamma \text{ is a maximal consistent set and if } [\pi]\psi \in \Sigma \text{ then } \psi \in \Gamma$$

I will prove in the appendix, as lemma A.1, that $w_\Sigma[\![\pi]\!]v$ iff there is a Γ such that $v = w_\Gamma$ and $\Sigma R_\pi \Gamma$. The relevant step in the proof of the truth lemma then runs as follows:

$$\begin{array}{lcl}
w_\Sigma \models [\pi]\psi & \Leftrightarrow & w_\Sigma[\![\pi]\!] \models \psi \\
 & \Leftrightarrow & \text{for each } \Gamma: \text{ if } \Sigma R_\pi \Gamma \text{ then } w_\Gamma \models \psi \text{ (by lemma A.1)} \\
 & \Leftrightarrow & \psi \in \Gamma \text{ for each } \Gamma \text{ such that } \Sigma R_\pi \Gamma \text{ (by induction hypothesis)} \\
 & \Leftrightarrow & [\pi]\psi \in \Sigma \text{ (by definition of } R_\pi)
\end{array}$$

5 Update Semantics

Update semantics, as it is presented in Veltman (1996), has been an important source of inspiration for this paper. It turns out that update semantics can be seen as a special case of the present approach: update semantics can be seen as describing the updates of an information state of a single agent who has fully introspective knowledge.

In update semantics, sentences are interpreted as functions that operate on information states. Information states are sets of classical possible worlds. The relevant definitions are the following:

Definition 5.1

- $\mathcal{L}^{US}$ is the language built up from a set of propositional variables $\mathcal{P}$ and the connectives $\neg, \wedge$ and a unary sentence operator *might* in the obvious way.[8]

- A classical information state s is a set of classical possible worlds, i.e. a set of assignments of truth-values to the propositional variables.

[7] In the terminology of definition 2.4, this model is the solution of Σ in the standard canonical model for the minimal modal logic K.

[8] In Veltman's paper the language is restricted to those sentences in which *might* occurs only as the outermost operator in a sentence.

- For each sentence $\phi \in \mathcal{L}^{US}$ and each classical information state s, the update of s with ϕ, $s[\phi]$, is defined as:

$$\begin{array}{rcl} s[p] & = & \{w \in s \mid w(p) = 1\} \\ s[\phi \wedge \psi] & = & s[\phi] \cap s[\psi] \\ s[\neg\phi] & = & s \setminus s[\phi] \\ s[\mathit{might}\ \phi] & = & s \text{ if } s[\phi] \neq \emptyset \\ & = & \emptyset \text{ otherwise} \end{array}$$

- A sentence ϕ is accepted in an information state s, written as $s \models \phi$, just in case $s[\phi]s$. An argument $\phi_1 \ldots \phi_n/\psi$ is valid, written as $\phi_1 \ldots \phi_n \models_{US} \psi$, iff for each s: $s[\phi_1]\ldots[\phi_n] \models \psi$.

There is a close correspondence between updates of information states in update semantics and consciously updating in possibilities in which agents have introspective information. More precisely, we can associate with each introspective possibility w and agent a a classical information state w^a, which consists of the set of classical worlds that correspond to the possibilities in $w(a)$. Vice versa, given a classical world w and an agent a, we can associate with each classical information state s a possibility s^a_w that assigns to each propositional variable the same value as w does and that assigns to a a set containing a possibility s^a_v for each $v \in s$ (a classical information state does not provide us with any information about which agent we are talking about, or what the 'real world' looks like, so we have to supply these parameters ourselves).

More formally:

Definition 5.2

- If w is a possibility and a an agent, then $w^a = \{v \text{ restricted to } \mathcal{P} \mid v \in w(a)\}$.
- If s is a classical information state, w a classical possible world, and a an agent, then s^a_w is a possibility such that $s^a_w(p) = w(p)$ for each $p \in \mathcal{P}$, and $s^a_w(a) = \{s^a_v \mid v \in s\}$.

It is not hard to see that s^a_w is an introspective possibility. The following proposition expresses how US-updates can be viewed as conscious updates of an introspective information state, if one reads $\neg\Box\neg$ for *might* . More precisely, seeing a classical information state as the information state of a certain agent a, updating such an information state with *might* ϕ in update semantics corresponds to updating a's information state with the test $?\neg\Box_a\neg\phi$.

Proposition 5.3 For each $\phi \in \mathcal{L}^{US}$, let ϕ' be just as ϕ but with all occurrences of *might* replaced by $\neg\Box_a\neg$. Then it holds that:

- For all classical information states s and t: $s[\phi]t$ iff $s^a_w[\![U^*_a ?\phi']\!]t^a_w$.

- For all possibilities w and v and each a such that a has introspective information in w: $w[\![U_a^*?\phi']\!]v$ iff $w^a[\phi]v^a$.
- $\phi_1 \ldots \phi_n \models_{US} \psi$ iff for all introspective w: $w \models [U_a^*?\phi_1'] \ldots [U_a^*?\phi_n']\Box_a\psi'$.

What this proposition expresses is that a US-update can be seen as a conscious update of the information state of an agent who has fully introspective information.

A Appendix

In this appendix, I will prove the lemma that was needed in the completeness proof in section 4. The lemma is the following:

Lemma A.1 Let π be a program, and assume that for each maximal consistent set of sentences Σ and each subprogram of the form $?\phi$ of π it holds that $w_\Sigma \models \phi$ iff $\phi \in \Sigma$. Then it holds for all Σ:

$$w_\Sigma[\![\pi]\!]v \text{ iff there is a } \Gamma \text{ such that } \Sigma R_\pi \Gamma \text{ and } v = w_\Gamma$$

proof: The proof is by induction on the structure of π.

- Tests: π is of the form $?\phi$.
  It follows by axiom 4 and maximality of Σ that $\Sigma R_{?\phi}\Sigma$ iff $\phi \in \Sigma$. The argument is then quite simple: $w_\Sigma[\![?\phi]\!]v$ iff $v = w_\Sigma$ and $w_\Sigma \models \phi$ iff $\phi \in \Sigma$ and $\Sigma R_{?\phi}\Sigma$.
- Conscious updates: $U_{\mathcal{B}}^*\pi$. This step is proven in lemma A.3.
- Disjunction: $\pi \cup \pi'$.
  Assume $w_\Sigma[\![\pi \cup \pi']\!]v$. Then $w_\Sigma[\![\pi]\!]v$ or $w_\Sigma[\![\pi']\!]v$. By induction hypothesis, there must be a Γ such that $v = w_\Gamma$ and $\Sigma R_\pi \Gamma$ or $\Sigma R_{\pi'}\Gamma$. To show that $\Sigma R_{\pi\cup\pi'}\Gamma$ take any $[\pi \cup \pi']\psi \in \Sigma$. Then $[\pi]\psi \wedge [\pi']\psi \in \Sigma$, and hence, by maximality of Σ, both $[\pi]\psi \in \Sigma$ and $[\pi']\psi \in \Sigma$. But that means, since $\Sigma R_\pi \Gamma$ or $\Sigma R_{\pi'}\Gamma$, that $\psi \in \Gamma$.
  For the other direction, assume that it does not hold that $w_\Sigma[\![\pi \cup \pi']\!]w_\Gamma$. Then neither $w_\Sigma[\![\pi]\!]w_\Gamma$, nor $w_\Sigma[\![\pi']\!]w_\Gamma$. So, by induction hypothesis, there is a ψ such that $[\pi]\psi \in \Sigma$ but $\psi \notin \Gamma$, and there is a ψ' such that $[\pi']\psi' \in \Sigma$ but $\psi' \notin \Gamma$. But then, $[\pi](\psi \vee \psi') \in \Sigma$ and $[\pi'](\psi \vee \psi') \in \Sigma$, whence $[\pi \cup \pi'](\psi \vee \psi') \in \Sigma$. But by maximality of Γ, $\neg\psi \wedge \neg\psi' \in \Gamma$, so it is not the case that $\Sigma R_{\pi\cup\pi'}\Gamma$.
- Sequencing: $\pi; \pi'$.
  Assume that $w_\Sigma[\![\pi; \pi']\!]v$. By induction hypothesis, there must be Δ and Γ such that $w_\Sigma[\![\pi]\!]w_\Delta[\![\pi']\!]w_\Gamma$, and $v = w_\Gamma$. Assume $[\pi; \pi']\psi \in \Sigma$. Then, by induction hypothesis, $[\pi']\psi \in \Delta$, and $\psi \in \Gamma$. Since $[\phi; \phi']\psi$ was arbitrary,

it follows that $\Sigma R_{\pi;\pi'}\Gamma$.

For the other direction, assume that $\Sigma R_{\pi;\pi'}\Gamma$. This means that the set $\{\psi \mid [\pi][\pi']\psi \in \Sigma\}$ is consistent (since it is a subset of Γ), and hence, $\{\psi \mid [\pi]\psi \in \Sigma\}$ is consistent. But then there is a Δ such that $\Sigma R_\pi \Delta$ and $\Delta R_{\pi'}\Gamma$. But then, by induction hypothesis, $w_\Sigma [\![\pi]\!] w_\Delta [\![\pi']\!] w_\Gamma$, and hence, $w_\Sigma [\![\pi;\pi']\!] w_\Gamma$.

Before giving the proof for the second step in the induction, I would like to make a general remark about the method of proof that will be used. We will prove that two possibilities are equal by showing that there exists a bisimulation between them. A relation $\mathcal{R}$ is a bisimulation between possibilities iff for every two possibilities w and v it holds that if $w\mathcal{R}v$, then (1) $w(p) = v(p)$ for each $p \in \mathcal{P}$, and (2) for each $w' \in w(a)$ there is a $v' \in v(a)$ such that $w'\mathcal{R}v'$, and (3) for each $v' \in v(a)$ there is a $w' \in w(a)$ such that $w'\mathcal{R}v'$. It turns out that the following proposition, which is closely related to proposition 2.5, is a consequence of the axiom of anti-foundation:

Proposition A.2 $w = v$ iff there is a bisimulation $\mathcal{R}$ such that $w\mathcal{R}v$.

I will make use of this fact in the proof of the following lemma.

Lemma A.3 Fix any $U^*_{\mathcal{B}}\pi$ and assume that it holds that $w_\Sigma [\![\pi]\!] v$ iff there is a Γ such that $\Sigma R_\pi \Gamma$ and $v = w_\Gamma$. (This is the induction hypothesis of the previous lemma.) It holds that:

$$w_\Sigma [\![U^*_{\mathcal{B}}\pi]\!] v \text{ iff there is a } \Gamma \text{ such that } v = w_\Gamma \text{ and } \Sigma R_{U^*_{\mathcal{B}}\pi}\Gamma$$

proof: Note that by axiom 5, the set $\{\psi \mid [U^*_{\mathcal{B}}\pi]\psi \in \Sigma\}$ is maximal consistent if Σ is. That means that there always is a unique Γ such that $\Sigma R_{U^*_{\mathcal{B}}\pi}\Gamma$. This implies that to prove the lemma, it is enough to show that if $w_\Sigma [\![U^*_{\mathcal{B}}\pi]\!] v$, and $\Sigma R_{U^*_{\mathcal{B}}\pi}\Gamma$, then there is a bisimulation between v and w_Γ. By proposition A.2 this shows that v and w_Γ are in fact equal.

Given a program π and a set of agents $\mathcal{B}$, define a relation $\mathcal{R}$ on possibilities by

$$\begin{array}{lll} w\mathcal{R}v & \text{iff} & w = v \text{ or there exist maximal consistent} \\ & & \text{sets } \Sigma \text{ and } \Gamma \text{ such that } w_\Sigma [\![U^*_{\mathcal{B}}\pi]\!] v, \Sigma R_{U^*_{\mathcal{B}}\pi}\Gamma \text{ and } w = w_\Gamma \end{array}$$

We will show that $\mathcal{R}$ is a bisimulation. Let $w\mathcal{R}v$, and let Σ and Γ be such that $w_\Sigma [\![U^*_{\mathcal{B}}\pi]\!] v$, $\Sigma R_{U^*_{\mathcal{B}}\pi}\Sigma$, and $w = w_\Gamma$ (the case that $w = v$ is easy). We need to show that the three clauses that define a bisimulation hold:

(1)

We first show that $w(p) = v(p)$ for all $p \in \mathcal{P}$.

$v(p) = 1$ iff $w_\Sigma [\![U^*_{\mathcal{B}}\pi]\!](p) = 1$ iff $w_\Sigma(p) = 1$ (by the semantics) iff $p \in \Sigma$ (by the definition of w_Σ) iff $p \in U^*_{\mathcal{B}}\pi(\Sigma)$ (by axiom 6) iff $w_\Gamma(p) = 1$.

(2)
Next we must show that for each $a \in \mathcal{A}$, if $v' \in v(a)$ then there is a $w' \in w_\Gamma(a)$ such that $w'\mathcal{R}v'$. We distinguish two cases: $a \in \mathcal{B}$, and $a \notin \mathcal{B}$.

First assume that $a \notin \mathcal{B}$. It follows by axiom 8 that $\Box_a\psi \in \Sigma$ iff $\Box_a\psi \in \Gamma$, which implies that $w_\Sigma(a) = w_\Gamma(a)$. But by the definition of $\llbracket U^*_{\mathcal{B}}\pi \rrbracket$ and the fact that $a \notin \mathcal{B}$, we have $w_\Sigma(a) = v(a)$. That means that $w_\Gamma(a) = v(a)$, which is sufficient, since by definition $\mathcal{R}$ includes the identity relation.

For the case that $a \in \mathcal{B}$, take any $v' \in v(a)$. We need to show that there is a $w_{\Gamma'} \in w_\Gamma(a)$ such that $w_{\Gamma'}\mathcal{R}v'$. The following picture might make matters more clear.

$$\begin{tikzcd}
w_\Gamma \arrow[dd, "a"] & w_\Sigma \arrow[l, "{[U^*_{\mathcal{B}}\pi]\psi\in\Sigma \Rightarrow \psi\in\Gamma}"'] \arrow[rr, "\llbracket U^*_{\mathcal{B}}\pi \rrbracket"'] \arrow[d, "a"] & & v \arrow[dd, "a"] \\
& w_{\Sigma'} \arrow[dr, "\llbracket \pi \rrbracket"] & & \\
w_{\Gamma'} & & w_\Delta \arrow[ll, "{[U^*_{\mathcal{B}}\pi]\psi\in\Delta \Rightarrow \psi\in\Gamma'}"] \arrow[r, "\llbracket U^*_{\mathcal{B}}\pi \rrbracket"] & v'
\end{tikzcd}$$

Since $v' \in v(a)$, there must be, by definition of $\llbracket U^*_{\mathcal{B}}\pi \rrbracket$, a Σ' such that $w_{\Sigma'} \in w_\Sigma(a)$ and a u such that $w_{\Sigma'}\llbracket\pi\rrbracket u\llbracket U^*_{\mathcal{B}}\pi\rrbracket v'$. By induction hypothesis, then, there is a Δ such that $\Sigma' R_\pi \Delta$ and $u = w_\Delta$. Take any such Δ and consider the set $\Gamma' = \{\psi \mid [U^*_{\mathcal{B}}\pi]\psi \in \Delta\}$. By the functionality axiom 5, this set is maximal consistent.. From the definition of Δ and Γ' it then follows that $w_\Delta\llbracket U^*_{\mathcal{B}}\pi\rrbracket v'$ and that $\Delta R_{U^*_{\mathcal{B}}\pi}\Gamma'$, and hence that $w_{\Gamma'}\mathcal{R}v'$.

To show that $w_{\Gamma'} \in w_\Gamma(a)$, take any $\Box_a\psi \in \Gamma$. Then $[U^*_{\mathcal{B}}\pi]\Box_a\psi \in \Sigma$ (by axiom 5), and by axiom 7, $\Box_a[\pi][U^*_{\mathcal{B}}\pi]\psi \in \Sigma$. But then, $[\pi][U^*_{\mathcal{B}}\pi]\psi \in \Sigma'$, so $[U^*_{\mathcal{B}}\pi]\psi \in \Delta$, and hence, $\psi \in \Gamma'$.

(3)
Finally we must show that for each $a \in \mathcal{A}$, if $w' \in w_\Gamma(a)$ then there is a $v' \in v(a)$ such that $w'\mathcal{R}v'$.

If $a \notin \mathcal{B}$, we can use the same argument as in case (2).

For the other case, take any Γ' such that $w_{\Gamma'} \in w(a)$, i.e. such that $\Box_a\psi \in \Gamma \Rightarrow \psi \in \Gamma'$. We need to find a $v' \in v(a)$ such that $w_{\Gamma'}\mathcal{R}v'$.

What we will do is show that there must exist maximal consistent sets Σ' and Δ such that: (i) $\Box_a\psi \in \Sigma \Rightarrow \psi \in \Sigma'$ (ii) $[\pi]\psi \in \Sigma' \Rightarrow \psi \in \Delta$ and (iii) $[U^*_{\mathcal{B}}\pi]\psi \in \Delta \Leftrightarrow \psi \in \Gamma'$.

From (i) it follows that $w_{\Sigma'} \in w_{\Sigma}(a)$ and from (ii) that $w_{\Sigma'}[\![\pi]\!]w_{\Delta}$. Since $[\![U^*_{\mathcal{B}}\pi]\!]$ is a total function, there must be a v' such that $w_{\Delta}[\![U^*_{\mathcal{B}}\pi]\!]v'$, and since $w_{\Sigma}[\![U^*_{\mathcal{B}}\pi]\!]v$, it follows that $v' \in v(a)$. Finally, it follows from (iii) that $\Delta R_{U^*_{\mathcal{B}}}\Gamma'$ and hence that $w_{\Gamma'}\mathcal{R}v'$. (See the picture above.)

To show the existence of the sets Σ' and Δ, consider first the following set σ (the notation $\langle\pi\rangle$ stands for $\neg[\pi]\neg$):

$$\{\psi \mid \Box_a\psi \in \Sigma\} \cup \{\langle\pi\rangle[U^*_{\mathcal{B}}\pi]\psi \mid \psi \in \Gamma'\} \qquad (\sigma)$$

We show that this set is consistent. For assume it is not. Then there must be $\phi_1 \ldots \phi_n$ such that $\Box_a\phi_i \in \Sigma$ for $i \leq n$ and $\psi_1 \ldots \psi_m$ such that $\psi_i \in \Gamma'$ for $i \leq m$ for which the following holds:

$\phi_1 \ldots \phi_n, \langle\pi\rangle[U^*_{\mathcal{B}}\pi]\psi_1 \ldots \langle\pi\rangle[U^*_{\mathcal{B}}\pi]\psi_m \vdash \bot$. That means that
$\phi_1 \ldots \phi_n \vdash [\pi]\neg[U^*_{\mathcal{B}}\pi]\psi_1 \vee \ldots \vee [\pi]\neg[U^*_{\mathcal{B}}\pi]\psi_m$, whence by axiom 5
$\phi_1 \ldots \phi_n \vdash [\pi][U^*_{\mathcal{B}}\pi]\neg(\psi_1 \wedge \ldots \wedge \psi_m)$, so, using the necessitation rule,
$\Box_a\phi_1 \ldots \Box_a\phi_n \vdash \Box_a[\pi][U^*_{\mathcal{B}}\pi]\neg(\psi_1 \wedge \ldots \wedge \psi_m)$, which means that
$\Sigma \vdash [U^*_{\mathcal{B}}\pi]\Box_a\neg(\psi_1 \wedge \ldots \wedge \psi_m)$. But then
$\Box_a\neg(\psi_1 \wedge \ldots \wedge \psi_m) \in \Gamma$, and thus
$\neg(\psi_1 \wedge \ldots \wedge \psi_m) \in \Gamma'$, in contradiction with the consistency of Γ' and the assumption that $\psi_i \in \Gamma$ for all $i \leq m$.

So, since the set σ is consistent, it has, by Lindenbaum's lemma, a maximal consistent extension Σ'. Take any such Σ', and consider the set δ:

$$\{\psi \mid [\pi]\psi \in \Sigma'\} \cup \{[U^*_{\mathcal{B}}\pi]\psi \mid \psi \in \Gamma'\} \qquad (\delta)$$

We show that δ is consistent as well, by a similar argument. For if it is not, there must be $\phi_1 \ldots \phi_n$ and $\psi_1 \ldots \psi_m$ such that $[\pi]\phi_i \in \Sigma'$ for $i \leq n$ and $\psi_i \in \Gamma'$ for each $i \leq m$ such that:

$\phi_1 \ldots \phi_n, [U^*_{\mathcal{B}}\pi]\psi_1 \ldots [U^*_{\mathcal{B}}\pi]\psi_m \vdash \bot$, so, using axiom 5,
$\phi_1 \ldots \phi_n \vdash \neg[U^*_{\mathcal{B}}\pi](\psi_1 \wedge \ldots \wedge \psi_m)$, and using necessitation
$[\pi]\phi_1 \ldots [\pi]\phi_n \vdash [\pi]\neg[U^*_{\mathcal{B}}\pi](\psi_1 \wedge \ldots \wedge \psi_m)$,
$[\pi]\phi_1 \ldots [\pi]\phi_n, \langle\pi\rangle[U^*_{\mathcal{B}}\pi](\psi_1 \wedge \ldots \wedge \psi_m) \vdash \bot$.

This contradicts the fact that Σ' is consistent, because, since $\psi_i \in \Gamma'$ for each $i \leq m$, it holds that $\psi_1 \wedge \ldots \wedge \psi_m \in \Gamma'$, and hence that $\langle\pi\rangle[U^*_{\mathcal{B}}\pi](\psi_1 \wedge \ldots \wedge \psi_m) \in \Sigma'$, while $[\pi]\phi_i \in \Sigma'$ for each $i \leq n$.

So, the set δ has a maximal consistent extension Δ, by Lindenbaum's lemma. By definition of Δ, properties (i), (ii) and (iii) hold, which completes the proof.

References

Peter Aczel. *Non-well-founded Sets.* CSLI Lecture Notes 14. Stanford, 1988.

Jon Barwise. On the model theory of common knowledge. In *The Situation in Logic*, CSLI Lecture Notes, pages 201–220. Stanford, 1989.

Jon Barwise and Lawrence S. Moss. *Vicious Circles.* CSLI Lecture Notes 60, Stanford, 1996.

David Beaver. *Presupposition and Assertion in Dynamic Semantics.* PhD thesis, Center for Cognitive Science, University of Edinburgh, 1995.

Johan van Benthem. *Exploring Logical Dynamics.* Studies in Logic, Language and Information. CSLI Publications, Stanford, 1996.

Maarten de Rijke. A system of dynamic modal logic. ILLC Research Report LP-92-08, 1992. To appear in *The Journal of Philosophical Logic.*

R. Fagin, J.Y. Halpern, Y. Moses, and M. Vardi. *Reasoning about Knowledge.* The MIT Press, Cambridge (Mass.), 1995.

Jelle Gerbrandy and Willem Groeneveld. Reasoning about information change. To appear in the *Journal of Logic, Language and Information.*

Robert Goldblatt. *Logics of Time and Computation.* CSLI Lecture Notes 7. Stanford, 1987.

Jeroen Groenendijk and Martin Stokhof. Dynamic predicate logic. *Linguistics and Philosophy*, 14(1):39–100, 1991.

Willem Groeneveld. *Logical Investigations into Dynamic Semantics.* PhD thesis, ILLC Dissertation Series 1995-18, 1995.

D. Harel. Dynamic logic. In D. M. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic, Vol. 2*, pages 497–604. Reidel, Dordrecht, 1984.

Jan Jaspars. *Calculi for Constructive Communication.* ILLC Dissertation Series 1994-4, ITK Dissertation Series 1994-1, 1994.

Robert C. Stalnaker. Pragmatics. In Harman and Davidson, editors, *Semantics of Natural Language*, pages 380–397. D. Reidel, 1972.

Frank Veltman. Defaults in update semantics. *Journal of Philosophical Logic*, 25:221–261, 1996.