Reasoning about Information Change

Jelle Gerbrandy and Willem Groeneveld

1 Introduction

Recently the notions of information and information change have gained a prominent place in several fields of scientific research, such as philosophy, the formal semantics of natural language and computer science. The present paper highlights two of such developments, namely the update semantics of Veltman (1996), and the analysis of communication in distributed systems, in particular the approach of Fagin et al. (1995). Our main goal here is to show that tools of the former may provide useful supplements to the approach of the latter.

2 Update Semantics

In his influential paper 'Defaults in Update Semantics,' Frank Veltman presents a dynamic view on meaning. Following the slogan "You know the meaning of a sentence if you know the change it brings about in the information state of anyone who accepts the news conveyed by it," the meaning of a sentence is associated with a function on information states. We will discuss Veltman's update semantics as a simple example to illustrate the main notions involved. The language of update semantics is a standard propositional modal language.

Definition 2.1 (Language) Given a set of propositional variables $\mathcal{P}$, the language $\mathcal{L}_\mathcal{P}$ of update semantics is the smallest set containing $\mathcal{P}$ such that if ϕ and ψ are in $\mathcal{L}_\mathcal{P}$, then $\phi \wedge \psi$, $\neg\phi$ and $\Diamond\phi$ are in $\mathcal{L}_\mathcal{P}$. □

As said, we will interpret sentences as functions on information states. The notion of information state used in update semantics is a very simple one.[1]

Definition 2.2 (Information states)

- A possible world w assigns to each propositional variable a truth-value; it is a function $w : \mathcal{P} \mapsto \{0, 1\}$.
- An information state σ is a set of possible worlds. □

[1] Veltman introduces more complex notions of information state to model reasoning with default rules.

Intuitively, if an agent is in information state σ, then σ contains all worlds that are compatible with the agent's information: for all the agent knows, each possible world in σ may picture reality correctly. For example, the information state consisting of the set of all possible worlds represents an information state of an agent that has no information about the world at all, and the empty set is an information state in which an agent has contradictory information.

When σ is an information state, and ϕ a sentence, we will write $\sigma[\phi]$ for the result of updating σ with the sentence ϕ. Intuitively, $\sigma[\phi]$ is the state that results when the agent, being in state σ, gets the information expressed by ϕ.

Definition 2.3 (Interpretation)

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

□

An update of an information state σ with a sentence p results in a state containing all and only the worlds in σ in which p is true. The result of updating a state σ with a negated sentence is a state containing all worlds in σ that do *not* survive in σ updated with ϕ. Conjunction is defined as intersection.

A sentence of the form $\Diamond\phi$ gets an interpretation roughly corresponding to the intuitive meaning of 'It might be the case that ϕ.' An update of σ with $\Diamond\phi$ either returns the same information state (when σ updated with ϕ does not result in the empty state, i.e. when ϕ is compatible with the information contained in σ), or it returns the inconsistent state (when ϕ is not compatible with the information contained in σ). This reflects the assumption that an agent in a state σ already knows what she considers possible and what not, which means that a sentence of the form 'It might be that...' can never provide new information; at most, it can be inconsistent with the information contained in σ.

Definition 2.4 (Acceptance and validity)

- A sentence ϕ is accepted in an information state σ iff $\sigma[\phi] = \sigma$. Notation; $\sigma \Vdash \phi$.
- A sequence of sentences $\phi_1 \ldots \phi_n$ is acceptable iff there is a σ such that $\sigma[\phi_1]\ldots[\phi_n] \neq \emptyset$.
- An argument $\phi_1 \ldots \phi_n/\psi$ is valid iff updating any information state with the premises in the order they are given, result in a state in which the conclusion is accepted: $\phi_1 \ldots \phi_n \models \psi$ iff for each σ: $\sigma[\phi_1]\ldots[\phi_n] \Vdash \psi$

A sentence is accepted in an information state if an update with ϕ does not change the information state. Intuitively, this happens only when the information that ϕ is already contained in σ. A sequence of sentences is acceptable when there is a state in which updating with the sentences in the order they are given does not result in the inconsistent empty state. An argument is valid iff updating an information state with the premises in the order they are given, results in a state in which the conclusion is accepted.

One of the interesting features of update semantics is that the order in which sentences in a text occur is important. In general, it is not the case that $\sigma[\phi][\psi] = \sigma[\psi][\phi]$. This correctly reflects the fact that changing the order of the sentences of a text will in general produce a different story, which need not even be coherent. Consider for example the following two examples:

Someone is knocking at the door...It might be John...It is Mary.

Someone is knocking at the door...It is Mary...It might be John.

The first sequence of sentences is acceptable, while the second is not: once one knows it is Mary who is knocking at the door, it cannot be John anymore. This is reflected in update semantics: $\Diamond p, \neg p$ is acceptable, while $\neg p, \Diamond p$ is not.

Another feature of update semantics is that updates always imply an increase of information, in the sense that an update of an information state σ always results in a state that is a subset of σ. i.e. $\sigma[\phi] \subseteq \sigma$. But this does not mean all sentences that are accepted in the first state are also accepted in the updated state. A typical case is a state σ containing two worlds w and v, with $w(p) = 1$ and $v(p) = 0$. This is a state in which the agent does not know whether p is true or not. In such a state, $\Diamond p$ is accepted. But as soon as the agent learns that p is not the case, p is not considered possible anymore: in the state $\sigma[\neg p]$, $\Diamond p$ is not accepted.

3 Dynamic Epistemic Semantics

Update semantics is a semantics that models the information change of a single agent. In this section, we develop a semantics for a language in which it is possible to express facts about the information and information change of several agents.[2] The language is the following:

Definition 3.1 (Language of DES)
Let $\mathcal{A}$ be a non-empty set of agents, and let $\mathcal{P}$ be a set of propositional atoms. The language $\mathcal{L}_{\mathcal{P}}^{\mathcal{A}}$ of DES is the smallest set such that $\mathcal{P} \subseteq \mathcal{L}$, and if ϕ and ψ are in $\mathcal{L}_{\mathcal{P}}^{\mathcal{A}}$, and $a \in \mathcal{A}$, then $\neg\phi, \phi \wedge \psi, \Box_a\phi$ and $[\phi]_a\psi$ are in $\mathcal{L}_{\mathcal{P}}^{\mathcal{A}}$.

[2] For an approach that uses a similar language, and employs a notion of constructive update over partial Kripke models, see Jaspars (1994).

Other logical constants, such as $\vee$, $\rightarrow$ and $\diamond_a$, are defined in the standard way. We will refer to the part of the language that does not contain the $[\cdot]_a$-operator as the 'classical fragment of the language'. This is just the language of classical multi-modal logic.

The intended interpretation of $\Box_a \phi$ is that agent a has the information that ϕ. The intended meaning of $[\phi]_a \psi$ is that an update of a's information with ϕ results in a situation where ψ is true. There is an operator $[\phi]_a$ for each agent a and each sentence ϕ in the language, which reflects the idea that any statement about the system of agents and their beliefs is something which the the agents may learn. This makes the agents effectively as 'intelligent' as us theoreticians, i.e. in principle any property of the system we are able to formulate in the object language may be known or learned by the agents.

To give a semantics for this language, we first need to make a choice as to how to represent the information of the agents. We believe that a representation that is based on non-well-founded sets is the most elegant way to do this.[3]

Definition 3.2 (Possibilities)
Let $\mathcal{A}$, a set of agents, and $\mathcal{P}$, a set of propositional variables, be given.

- 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. □

Clearly, this definition is circular, since possibilities are defined in terms of information states, and information states are sets of possibilities. In the universe of non-well-founded sets of Aczel (1988), this circularity is harmless.[4]

A possibility w characterizes which propositions are true and which are false by assigning to each atomic sentence a truth value, and it characterizes the information of each of the agents by assigning to each agent an information state. The information of an agent is represented, as it is in update semantics, as a set of possible ways the world might be, according to that agent. In this case, this is a set of possibilities.

There is a close relation between these non-well-founded models and Kripke-structures; to be precise, there is a one-one-relation between possibilities and bisimulation classes of worlds in Kripke models, that preserves truth of the classical fragment of the language.

Definition 3.3 Let $M = (W, (R_a)_{a \in A}, V)$ be a Kripke model.
A *decoration* of M is a function that assigns to each $w \in W$ a function $d(w)$ on $\mathcal{P} \cup \mathcal{A}$ such that

[3] See Groeneveld (1995) for a discussion of the semantics of DES using Kripke models and the knowledge structures of Fagin and Halpern (Fagin et al. 1991).

[4] The underlying set-theory is axiomatized by ZFC^- (the Zermelo-Fraenkel axioms minus the axiom of foundation) plus Aczel's Anti-Foundation Axiom (AFA).

- $d(w)(p) = V(w)(p)$, i.e. $d(w)$ assigns to each propositional variable the same truth-value as it has in w in the model.
- $d(w)(a) = \{d(v) \mid wR_a v\}$, i.e. $d(w)$ assigns to each agent a the set of functions associated with worlds reachable from w by R_a.

If d is a decoration of M, and w a world in M, we say that $d(w)$ is the solution of w in M, and (M, w) is a picture of $d(w)$. □

Proposition 3.4

- Each Kripke model M has a unique decoration. This decoration assigns to each world in M a possibility.
- Each possibility has a picture.
- w in M and w' in M' have the same solution iff w and w' are bisimilar. □

This means that possibilities can be seen as representing bisimulation classes of worlds in Kripke models. Moreover, it implies the Bisimulation Principle (3.5 below) which we will frequently use later. In the following definition, we use the notation $w[\mathcal{B}]v$, for $\mathcal{B}$ a set of agent, to stand for the fact that w and v differ at most from each other in the information states they assign to agent in $\mathcal{B}$.

Proposition 3.5 (Bisimulation Principle) A bisimulation between possibilities is any relation B such that wBv iff $w[\mathcal{A}]v$ and for each $a \in \mathcal{A}$, if $w' \in w(a)$, then there is a $v' \in v(a)$ such that $w'Bv'$, and if $v' \in v(a)$, then there is a $w' \in w(a)$ such that $w'Bv'$.

If B is a bisimulation, then for all possibilities w, v: if wBv then $w = v$ □

Properties of possibilities

One of the charms of Kripke semantics is the fact that properties of information such as positive introspection or consistency correspond to certain simple properties on frames, such as transitivity and seriality of the accessibility relations. Here are some examples of constraints on possibilities that correspond to familiar frame constraints.

Definition 3.6 Call a class of possibilities S closed iff it holds that if $w \in S$ and $v \in w(a)$ then $v \in S$.

1. $\mathcal{C}$, the class of consistent possibilities is the largest closed class such that $w \in \mathcal{C}$ implies $w(a) \neq \emptyset$
2. $\mathcal{T}$, the class of truthful possibilities is the largest closed class such that $w \in \mathcal{T}$ implies $w \in w(a)$

3. $\mathcal{P}$, the class of positive introspective possibilities is the largest closed class such that $w \in \mathcal{P}$ and $v \in w(a)$ imply $v(a) \subseteq w(a)$

4. $\mathcal{N}$, the class of negative introspective possibilities is the largest closed class such that $w \in \mathcal{N}$ and $v \in w(a)$ imply $w(a) \subseteq v(a)$ $\square$

Of special interest is the class of fully introspective possibilities $\mathcal{P} \cap \mathcal{N}$ (which is a closed class).

Conscious updates

In update semantics, if an agent updates her information with p she will discard all possible worlds in which p is false. But if her epistemic alternatives are not classical possible worlds, but possibilities as we have defined them, there will be no point in also preserving those options in which p is true but in which the agent does not have the information that p. Ideally, she will also accommodate for the fact that she has learned p, and after having learned that p, she will not only have the information that p, but on top of that have the information that she has the information that p. And so on. We will refer to such an update as a 'conscious' update: the agent who gets new information is conscious of the fact that she gets this new information.

Note that this is not an 'eliminative' process. A conscious update is not one in which one simply discards possibilities to reach a state in which one has more information. For example, consider a situation in which an agent a does not know whether p, and knows that she does not know this. This will be modeled by a possibility w such that the set of possibilities $w(a)$ will only contain possibilities in which a does not know whether p. Removing all non-p-possibilities from this information state leaves the agent with a set of possibilities in which p is true, but in which she does not know whether p.

We use the idea of conscious update for for interpreting sentences of the form $[\phi]_a\psi$, which will be interpreted as 'after a consciously updates with ϕ, ψ is true.' To do this, we first have to give a formal definition of conscious update. We will define for each sentence ϕ and each $a \in \mathcal{A}$ a function $[\![\phi]\!]_a$ on possibilities in such a way that applying this function to a possibility returns a new possibility that is the result of updating a's information state consciously with the information that ϕ. Remember that we use the notation $w[a]v$ as an abbreviation for the statement that w and v differ at most in the information state they assign to a.

Definition 3.7 (Conscious updates)
$w[\![\phi]\!]_a$ is that w' such that $w[a]w'$ and $w'(a) = \{v[\![\phi]\!]_a \mid v \in w(a) \text{ and } v \models \phi\}$.

So a conscious update of a's information in a possibility w changes the possibility w in such a way that only a's information state is changed (i.e. the new possibility differs from the old one only in the information state that is assigned to a), and this is done in the following way: all possibilities in which ϕ is not true are eliminated from a's information state, and in all remaining possibilities, a's information state is consciously updated with ϕ.

That this notion of update is well-defined needs some proof. We will rely on the Solution Lemma of Aczel (1988), which is standardly used in the set theory ZFC/AFA for establishing the existence of non-well-founded sets, to prove that that there is in fact a relation that conforms to definition 3.7. Then, we will give an argument using the bisimulation principle to show that this relation is the only relation conforming to the definition.

To prove the existence of the update function, fix an actor $a \in \mathcal{A}$ and some proposition p (i.e. p is a class of possibilities). For each possibility w, introduce an indeterminate x_w, and consider the class of equations defined by the stipulations

$$x_w = \{(q,i) \mid w(q) = i\} \cup \{(b, w(b)) \mid b \neq a\} \cup \{(a, \{x_v \mid v \in w(a) \cap p\})\}$$

By the Solution Lemma this system has a unique solution, which in this case is a map π from indeterminates to possibilities. Then define $cu(a,p)$, the conscious update with p for a by

$$(w,v) \in cu(a,p) \text{ iff } v = \pi(x_w)$$

It is not hard to check that it holds that $(w,v) \in cu(a,p)$ iff $w[a]v$ and $v(a) = \{v' \mid \exists w' \in w(a) \cap p : (w',v') \in cu(a,p)\}$, which shows that $cu(a,p)$ is the function we were looking for.

To show that the update function is unique for each ϕ and a, assume that there are in fact two functions f and f' that conform to definition 3.7. Note that by the definition, both of these functions will be total on the class of all possibilities. We will show that it holds for each possibility w, that $f(w) = f'(w)$. For define a relation B as:

$$wBv \text{ iff } w = v \text{ or } \exists u : f(u) = w \text{ and } f'(u) = v$$

We claim that B is a bisimulation, from which it follows by the bisimulation principle that $w = v$. Clearly, $w[a]v$, so we need to show that for each $w' \in w(a)$ there is a $v' \in v(a)$ such that $w'Bv'$, and vice versa. Take any $w' \in w(a)$. Then, there must be a $u' \in u(a)$ such that $u' \models \phi$ and $f(u') = w'$. But then, since f' is a total function, there must be a $v' \in v(a)$ such that $f'(u') = v'$. But for this v' it holds that $w'Bv'$. The other direction is completely symmetric, which shows that B is a bisimulation.

Definition 3.8 (Truth)

$$\begin{array}{rcl} w \models p & \text{iff} & w(p) = 1 \\ w \models \phi \wedge \psi & \text{iff} & w \models \phi \text{ and } w \models \psi \end{array}$$

$$
\begin{array}{rcl}
w \models \neg\phi & \text{iff} & w \not\models \phi \\
w \models \Box_a \phi & \text{iff} & \forall v \in w(a) : v \models \phi \\
w \models [\phi]_a \psi & \text{iff} & w[\![\phi]\!]_a \models \psi
\end{array}
$$

Technically we have to conceive of definitions 3.7 and 3.8 as one simultaneous definition, since the definition of update uses the notion of truth and vice versa. This offers no problems, however, and we have only separated the two for clarity.

All classical logical operators are interpreted classically: a conjunction is true just in case both conjuncts are, a negation is true iff the negated sentence is not true, $\Box_a \phi$ is true just in case ϕ is true in each possibility in a's information state. New is the definition for $[\phi]_a \psi$: such a sentence is true in a possibility w exactly when ψ is true in the possibility that results from updating a's information state in w with ϕ. We define validity in the standard way, i.e. $\models \phi$ iff $w \models \phi$ for each possibility w, and for each set of sentences Γ, $\Gamma \models \phi$ iff for each possibility w such that $w \models \psi$ for each $\psi \in \Gamma$, $w \models \phi$.

Update Semantics

It turns out that validity in update semantics can be expressed in DES by identifying a US-update with a conscious update in DES in a fully introspective possibility. The validity of an argument—updating with the premises results in an information state in which the conclusion is accepted—can then be expressed in DES by a sentence expressing that after an agent consciously updates with the premises, then she will accept the conclusion, i.e. she will have the information that the conclusion holds.

Proposition 3.9 Let for each ϕ in the language of update semantics, ϕ' be just like ϕ, except that each occurrence of $\Diamond$ is replaced by $\Diamond_a$. Then:
$\phi_1 \ldots \phi_n \models_{US} \psi$ iff for all fully introspective w, $w \models [\phi_1]_a \ldots [\phi]_a \Box_a \psi$ $\square$

Group Updates

Common knowledge is a concept that crops up in several places in the literature on distributed systems in computer science, in the literature on game theory in philosophy and economics and in the literature on pragmatics in linguistics. The concept is most easily explained as follows: a sentence ϕ is common knowledge between a group of agents $\mathcal{B}$ just in case each agent in $\mathcal{B}$ knows ϕ, each agent knows of each other agent that he knows ϕ, and so on.

What we will do here is model the effect of a sentence *becoming* common knowledge between a certain group of agents, and add operators $[\phi]_{\mathcal{B}}$ for each $\mathcal{B} \subseteq \mathcal{A}$ to express this in the object language. Such an operator may be useful, for example, to formalize an idea in the theory of discourse that the purpose of an assertion is to extend the common knowledge of speaker and hearer.

Definition 3.10 (Conscious group update)
$w[\![\phi]\!]_{\mathcal{B}}$ is that w' such that $w[\mathcal{B}]w'$ and $w'(a) = \{v[\![\phi]\!]_{\mathcal{B}} \mid v \in w(a) \text{ and } v \models \phi\}$ for each $a \in \mathcal{B}$.

A group update with ϕ in a possibility w results in a new possibility in which only the information of the agents in the group has changed. For each agent in the group, the new information state consists of all the old possibilities in which ϕ is true updated with the information that ϕ becomes common knowledge between the agents in the group.

This definition is structurally similar to the definition given for conscious updates above. In fact, it holds that for each w, $w[\![\phi]\!]_a = w[\![\phi]\!]_{\{a\}}$. Also, the proof that definition 3.10 is in fact correct is entirely analogous to the coinduction argument we gave for definition 3.7, for which reason we won't repeat the argument.

We can now extend the truth definition of DES with the following clause:

$$w \models [\phi]_{\mathcal{B}}\psi \quad \text{iff} \quad w[\![\phi]\!]_{\mathcal{B}} \models \psi$$

Axiomatization

In this section, we provide an axiomatization of the language of DES with group updates, and prove that it is sound and complete with respect to the semantics.

Definition 3.11 (Conscious K)
The system CK is defined by the following axioms and rules.

Axioms

- **A1** $\vdash \phi$, if ϕ is valid in classical propositional logic
- **A2** $\vdash \Box_a(\phi \to \psi) \to (\Box_a\phi \to \Box_a\psi)$
- **A3** $\vdash [\chi]_{\mathcal{B}}(\phi \to \psi) \to ([\chi]_{\mathcal{B}}\phi \to [\chi]_{\mathcal{B}}\psi)$ (normality)
- **A4** $\vdash \neg[\phi]_{\mathcal{B}}\psi \leftrightarrow [\phi]_{\mathcal{B}}\neg\psi$ (functionality)
- **A5** $\vdash p \leftrightarrow [\phi]_{\mathcal{B}}p$, if p is an atom. (independence)
- **A6** $\vdash [\phi]_{\mathcal{B}}\Box_a\psi \leftrightarrow \Box_a(\phi \to [\phi]_{\mathcal{B}}\psi)$ if $a \in \mathcal{B}$ (Generalized Ramsey Axiom)
- **A7** $\vdash \Box_a\phi \leftrightarrow [\psi]_{\mathcal{B}}\Box_a\phi$ if $a \notin \mathcal{B}$. (Privacy Axiom)

Rules

- **MP** $\phi, \phi \to \psi \vdash \phi$
- **Nec□** If $\vdash \phi$ then $\vdash \Box_a\phi$
- **Nec[·]** If $\vdash \phi$ then $\vdash [\psi]_{\mathcal{B}}\phi$

$\Gamma \vdash_{CK} \phi$ iff there is a derivation of ϕ from assumptions in Γ. □

So, in addition to the rules and axioms of classical model logic, the deduction system consists axioms describing the interaction between the dynamic operators and the classical logical constants. Axiom 3 together with the rule Nec[$\cdot$] guarantee that the dynamic operators behave as normal modal operators. Axiom 4 expresses that updates are functional: if it is not the case that a certain sentence is true after an update with a certain sentence, then, since the update always gives a unique result, it must be the case that the negation of that sentence is true in the updated possibility. Axiom 5 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 6 expresses that if it is the case that after a group update with ϕ, some agent in the group knows that ψ, then that agent already knew that if ϕ were true, then after a group update with ϕ, ψ would be true, and vice versa. Axiom 7, finally, expresses that a group update has no effect on the information of agents outside of that group.

Proposition 3.12 (Soundness) If $\Gamma \vdash_{CK} \phi$ then $\Gamma \models \phi$.

proof: A standard induction; by way of illustration, we show the correctness of axiom 6, and leave the remaining cases to the reader. We have the following equivalences, if $a \in \mathcal{B}$:

$$\begin{array}{rcl} w \models [\phi]_{\mathcal{B}}\Box_a\psi & \Leftrightarrow & w[\![\phi]\!]_{\mathcal{B}} \models \Box_a\psi \\ & \Leftrightarrow & \forall v \in w[\![\phi]\!]_{\mathcal{B}}(a):\ v \models \psi \\ & \Leftrightarrow & \forall v : [\exists u \in w(a) : u \models \phi \ \&\ v = u[\![\phi]\!]_{\mathcal{B}} \Rightarrow v \models \psi] \\ & \Leftrightarrow & \forall u \in w(a):\ \text{if } u \models \phi \text{ then } u[\![\phi]\!]_{\mathcal{B}} \models \psi \\ & \Leftrightarrow & \forall u \in w(a):\ \text{if } u \models \phi \text{ then } u \models [\phi]_{\mathcal{B}}\psi \\ & \Leftrightarrow & w \models \Box_a(\phi \to [\phi]_{\mathcal{B}}\psi) \end{array}$$

Proposition 3.13 (Completeness) $\Gamma \vdash \phi$ iff $\Gamma \models \phi$.

proof: We use a variation on the classical Henkin proof for completeness of modal logic, showing that for each consistent set of sentences there is a possibility in which these sentences are true.

It is easy to show that each consistent set can be extended to a maximal consistent set (we will refer to this as 'Lindenbaum's Lemma'). 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 b: $w_\Sigma(b) = \{w_\Gamma \mid \Gamma$ is maximal consistent and if $\Box_b\psi \in \Sigma$, then $\psi \in \Gamma\}$.[5] We prove the usual Truth Lemma, that is, for each sentence χ it holds that $\chi \in \Sigma$ iff $w_\Sigma \models \chi$. Completeness then follows by standard argumentation. The Truth Lemma is proven by an induction on the structure of χ, in which all cases are standard, except the case where χ is of the form

[5] In the terminology of definition 3.3, this model is the solution of the standard canonical model for the minimal modal logic K. Alternatively, the existence of the possibilities w_Γ is easily proven by defining the appropriate set of equations, and an appeal to the Solution Lemma.

$[\phi]_{\mathcal{B}}\psi$. 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 $\mathcal{B}$-update of w_Σ with ϕ depends on the formulae of the form $[\phi]_{\mathcal{B}}\psi$ in Σ. This is reflected by the following operation on maximal consistent sets:

$$\Gamma \bullet_{\mathcal{B}} \phi =_{df} \{\psi \mid [\phi]_{\mathcal{B}}\psi \in \Gamma\}$$

The functionality axiom A4 will ensure that $\Gamma \bullet_{\mathcal{B}} \phi$ will be a maximal consistent set whenever Γ is. The step in the Truth Lemma for formulae of the form $[\phi]_{\mathcal{B}}\psi$ then proceeds as follows:

$$\begin{array}{rcl} w_\Sigma \models [\phi]_{\mathcal{B}}\psi & \Leftrightarrow & w_\Sigma[\![\phi]\!]_{\mathcal{B}} \models \psi \\ & \Leftrightarrow & w_{\Sigma\bullet_{\mathcal{B}}\phi} \models \psi \text{ (by 3.14 and induction hypothesis on } \phi) \\ & \Leftrightarrow & \psi \in \Sigma \bullet_{\mathcal{B}} \phi \text{ (by induction hypothesis on } \psi) \\ & \Leftrightarrow & [\phi]_{\mathcal{B}}\psi \in \Sigma \text{ (by definition of } \Sigma \bullet_{\mathcal{B}} \phi) \end{array}$$

Which only leaves the next lemma. □

Lemma 3.14 Let ϕ be fixed, and assume that for each maximal consistent Σ, $w_\Sigma \models \phi$ iff $\phi \in \Sigma$. Then for all Σ, $w_\Sigma[\![\phi]\!]_{\mathcal{B}} = w_{\Sigma\bullet_{\mathcal{B}}\phi}$

proof: Define a relation $\mathcal{R}$ on possibilities by

$$\begin{array}{rl} w\mathcal{R}v \Leftrightarrow & w = v \text{ or there exists a maximal consistent} \\ & \text{set } \Sigma \text{ such that } w = w_\Sigma[\![\phi]\!]_{\mathcal{B}} \text{ and } v = w_{\Sigma\bullet_{\mathcal{B}}\phi} \end{array}$$

We will show that $\mathcal{R}$ is a bisimulation. The Bisimulation Principle 3.5 then implies that $\mathcal{R}$ actually is an identity relation, which proves the lemma.

Let $w\mathcal{R}v$, and suppose $w = w_\Sigma[\![\phi]\!]_{\mathcal{B}}$ and $v = w_{\Sigma\bullet_{\mathcal{B}}\phi}$ (the case that $w = v$ is easy). We need to show three things:

1. $w[\mathcal{A}]v$.
   $w_\Sigma[\![\phi]\!]_{\mathcal{B}}(p) = 1$ iff $w_\Sigma(p) = 1$ (by the semantics) iff $p \in \Sigma$ (by the definition of w_Σ) iff $p \in \Sigma \bullet_{\mathcal{B}} \phi$ (by axiom 5) iff $w_{\Sigma\bullet_{\mathcal{B}}\phi}(p) = 1$.

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

   Assume that $b \notin \mathcal{B}$. It then follows by axiom 7 that $\Box_b\psi \in \Sigma$ iff $\Box_b\psi \in \Sigma \bullet_{\mathcal{B}} \phi$, which implies that $w_\Sigma(b) = w_{\Sigma\bullet_{\mathcal{B}}\phi}(b)$. But by the definition of $[\![\phi]\!]_{\mathcal{B}}$ and the fact that $b \notin \mathcal{B}$, we have $w_\Sigma(b) = w_\Sigma[\![\phi]\!]_{\mathcal{B}}(b)$, so actually $w(b) = v(b)$ in this case, which is certainly sufficient.

   For the other case, let $b \in \mathcal{B}$ and take any $w' \in w_\Sigma[\![\phi]\!]_{\mathcal{B}}(b)$. Then there must be a Σ' that is maximal consistent such that $w_{\Sigma'} \models \phi$, $w' = w_{\Sigma'}[\![\phi]\!]_{\mathcal{B}}$, and $w_{\Sigma'} \in w_\Sigma(b)$. The latter implies that if $\Box_b\chi \in \Sigma$, then $\chi \in \Sigma'$.

   We want to find a v' such that (I) $v' \in w_{\Sigma\bullet_{\mathcal{B}}\phi}(b)$ and (II) $w'\mathcal{R}v'$. Consider the set $\Sigma' \bullet_{\mathcal{B}} \phi$ and take $v' = w_{\Sigma'\bullet_{\mathcal{B}}\phi}$. It is then immediate that (II) holds. To see that (I) holds, take any $\Box_b\psi \in \Sigma \bullet_{\mathcal{B}} \phi$. Then $[\phi]_{\mathcal{B}}\Box_b\psi \in \Sigma$, and hence, by axiom 6, $\Box_b(\phi \to [\phi]_{\mathcal{B}}\psi) \in \Sigma$, and thus, $\phi \to [\phi]_{\mathcal{B}}\psi \in \Sigma'$. Because $w_{\Sigma'} \models \phi$, it follows by the main assumption of this lemma that $\phi \in \Sigma'$, and hence $[\phi]_{\mathcal{B}}\psi \in \Sigma'$, which means that $\psi \in \Sigma' \bullet_{\mathcal{B}} \phi$. Since $\Box_b\psi$ was arbitrary, it follows that $w_{\Sigma' \bullet_{\mathcal{B}} \phi} \in w_{\Sigma \bullet_{\mathcal{B}} \phi}(b)$, which is what we wanted to prove.

3. Finally we must show that for each $b \in \mathcal{A}$, if $v' \in v(b)$ then there is a $w' \in w(b)$ such that $w'\mathcal{R}v'$.
   If $b \notin \mathcal{B}$, we can use the same argument as in case (2). So assume instead that $b \in \mathcal{B}$, and take any $w_\Gamma \in w_{\Sigma \bullet_{\mathcal{B}} \phi}(b)$.

   Consider the set $\Delta = \{[\phi]_{\mathcal{B}}\chi \mid \chi \in \Gamma\} \cup \{\phi\} \cup \{\psi \mid \Box_b\psi \in \Sigma\}$. Below we show that Δ is consistent (III). Then Δ can be extended by Lindenbaum's Lemma to a maximal consistent set Σ', for which it holds, by definition of Δ that $w_{\Sigma'} \in w_\Sigma(b)$, $w_{\Sigma'} \models \phi$. This implies that $w_{\Sigma'}[\![\phi]\!]_{\mathcal{B}} \in w_\Sigma[\![\phi]\!]_{\mathcal{B}}(b)$ $(= w(b))$. Moreover, it is not hard to see that by definition of Δ and the functionality axiom it holds that $\Gamma = \Sigma' \bullet_{\mathcal{B}} \phi$. It then follows that that $w_{\Sigma'}[\![\phi]\!]_{\mathcal{B}}\mathcal{R}w_\Gamma$.

   To see (III), suppose to the contrary that Δ is inconsistent. Then there must be $\Box_b\psi_1, \ldots, \Box_b\psi_n \in \Sigma$ and $\chi_1, \ldots, \chi_m \in \Gamma$ such that:

   $\phi, \psi_1, \ldots, \psi_n, [\phi]_{\mathcal{B}}\chi_1, \ldots, [\phi]_{\mathcal{B}}\chi_m \vdash \bot$. That means that:

   $\psi_1, \ldots, \psi_n \vdash \phi \to \neg([\phi]_{\mathcal{B}}\chi_1 \wedge \cdots [\phi]_{\mathcal{B}}\chi_m)$, so (using A3 and A4)

   $\psi_1, \ldots, \psi_n \vdash \phi \to [\phi]_{\mathcal{B}}\neg(\chi_1 \wedge \cdots \wedge \chi_m)$, hence (by nec$\Box$ and axiom 2)

   $\Box_b\psi_1, \ldots, \Box_b\psi_n \vdash \Box_b(\phi \to [\phi]_{\mathcal{B}}\neg(\chi_1 \wedge \cdots \wedge \chi_m))$, so (by axiom 6)

   $\Box_b\psi_1, \ldots, \Box_b\psi_n \vdash [\phi]_{\mathcal{B}}\Box_b\neg(\chi_1 \wedge \cdots \wedge \chi_m)$, so

   $\Sigma \vdash [\phi]_{\mathcal{B}}\Box_b\neg(\chi_1 \wedge \cdots \wedge \chi_m)$, but then

   $\Box_b\neg(\chi_1 \wedge \cdots \wedge \chi_m) \in \Sigma \bullet_{\mathcal{B}} \phi$, which means that

   $\neg(\chi_1 \wedge \cdots \wedge \chi_m) \in \Gamma$, by the fact that $w_\Gamma \in w_{\Sigma \bullet_{\mathcal{B}} \phi}(b)$,

   contradicting the fact that Γ is consistent.

$\Box$

Some observations and remarks

Here follows a short list of validities and non-validities.

Proposition 3.15

1. $\vdash [\phi \wedge \psi]_{\mathcal{B}}\chi \leftrightarrow [\psi \wedge \phi]_{\mathcal{B}}\chi$ (there is no difference between updating with $\phi \wedge \psi$ and updating with $\psi \wedge \phi$)

2. $\nvdash [\phi]_{\mathcal{B}}[\psi]_{\mathcal{B}}\chi \leftrightarrow [\psi]_{\mathcal{B}}[\phi]_{\mathcal{B}}\chi$ (updating first with ϕ and then with ψ is different from updating with ψ first, and after that with ϕ.)

3. $\vdash [\phi]_{\mathcal{B}}[\psi]_{\mathcal{C}}\chi \leftrightarrow [\psi]_{\mathcal{C}}[\phi]_{\mathcal{B}}\chi$ if $\mathcal{B} \cap \mathcal{C} = \emptyset$ (updates of the information of one group does not affect updates of the information of a disjoint group)

4. If $\vdash \phi \leftrightarrow \psi$, then $\vdash [\phi]_{\mathcal{B}}\chi \leftrightarrow [\psi]_{\mathcal{B}}\chi$ (if two sentences have the same truth conditions, they are also equivalent as updates)

It is possible to combine the notion of conscious update with stronger epistemic logics. This is unproblematic for the logic of introspection: if we add the introspection axioms of K45 to the axioms of CK, the resulting logic CK45 is sound and complete with respect to the class of introspective possibilities. For other well known epistemic logics such as KD45 or S5, updates will be partial functions, since conscious updates don't necessarily preserve the properties of consistency and correctness.

We sketch some details for one special case, S5. We take K to be the class of truthful and introspective possibilities (positive and negative; see definition 3.6). The K-update $[\![\phi]\!]_a^K$ is then defined as the restriction of $[\![\phi]\!]_a$ to possibilities in K. This will make the updates partial functions. For example, for an atom p this will have the effect that an agent a can only learn p in a possibility w if $w \models p$. As for the effect on the axiomatics, we conjecture that the following is complete: add the axioms of S5 to CK; weaken the functionality axiom to $\neg[\phi]_a\psi \rightarrow [\phi]_a\neg\psi$; and compensate the loss of the existential part of the functionality axiom by adding the axiom $\langle\phi\rangle_a\top \leftrightarrow \phi$.

We end this brief discussion by a comment on our choice of operators. We could have added operators of the form $\mathcal{C}_{\mathcal{B}}$ to the language, one for each (non-empty) set of agents $\mathcal{B}$, to express the static concept of ϕ being common knowledge between the agents in $\mathcal{B}$. Transferring the definition of Fagin et al. (1995) to our framework, its definition could be the following:

$$w \models \mathcal{C}_{\mathcal{B}}\phi \quad \text{iff} \quad w \models \Box_{a_1} \ldots \Box_{a_n}\phi \text{ for each } \{a_1 \ldots a_n\} \subseteq \mathcal{B}$$

The reason we have not added these operators is that we have not yet found an axiomatization for the language with these operators. We hope to correct this omission in the near future.

4 An application: Automated Dirty Children

We have developed a language and a semantics for reasoning about information and information change of several agents, which in turn can reason about their own information and the information of other agents. We have provided an axiomatization for this semantics. In this section, we want to show that the logic developed above can be useful as a tool for analyzing problems concerning reasoning about information in a multi-agent setting. To show this, we consider a textbook case, the puzzle of the dirty children. This puzzle occurs under different guises in the literature: it is a variant of the puzzle of the cheating husbands (see for example Moses et al., 1986), the wise men puzzle (in e.g. McCarthy, 1990) and the the Conway-paradox (e.g. van Emde Boaz et al., 1980).

The puzzle

The description of the dirty children puzzle that we give here is adapted from Barwise (1981).

There are n children playing together. During their play some of the children, say k of them, get mud on their foreheads. Each can see the mud on others but not on his own forehead. Along comes a father, who says, "At least one of you has mud on your head." He then asks the following question, over and over: "Can any of you prove that you have mud on your head?" Assuming that all the children are perceptive, intelligent, truthful, and that they answer simultaneously, what will happen?

There is a "proof" that the first $k-1$ times the father asks the question, the children will all say "no" but that the k-th time the children that are dirty will answer "yes."

The proof is by induction on the number of dirty children k. For $k = 1$ the result is obvious: the dirty child sees that no one else is muddy, so he must be the muddy one. If there are two dirty children, say a and b, each answers "no" the first time, because of the mud on the other. But, when b says "no," a realizes he must be muddy, for otherwise b would have known the mud was on his head and answered "yes" the first time. Thus a answers "yes" the second time. b goes through the same reasoning. Now suppose there are three dirty children, a, b, c. Child a argues as follows. Assume I don't have mud on my head. Then, by the $k = 2$ case, both b and c will answer "yes" the second time. When they don't, he realizes that the assumption was false, that he *is* muddy, and so will answer "yes" on the third question. Similarly for b and c.

Formalization

We will show how one can formalize the description of the puzzle, and the reasoning involved, in dynamic epistemic semantics. The result, that after $m-1$ answers to the father's question, the children that are dirty know that they are dirty, will then be a theorem in the logic.

Let $\mathcal{A}$ be a set of children playing in the mud. Consider a language that contains a propositional atom p_a for each $a \in \mathcal{A}$, which we will take to express that child a is dirty. We start by introducing some convenient abbreviations:

- Each child can see the forehead of each of the other children. So, if a child is dirty, each of the other children knows that she is dirty. This can be expressed by the conjunction of all sentences of the form $(p_a \rightarrow \Box_b p_a) \wedge (\neg p_a \rightarrow \Box_b \neg p_a))$ for each a and b in $\mathcal{A}$ such that $a \neq b$. We abbreviate this conjunction by **vision** .

- It is common knowledge between all children that each forehead can be seen by each of the other children, i.e. **vision** is common knowledge between all children. We can express this as $C_{\mathcal{A}}\mathbf{vision}$.

- Before asking the children whether they know if they are dirty or not, the father announces in face of all children that at least one of them has a dirty forehead. Let **father** be the sentence $\bigvee\{p_a \mid a \in \mathcal{A}\}$, which expresses that at least one of the children is dirty.

- After the father's announcement, all children answer the question 'Do you know whether you are dirty or not?' The children answer either 'yes' or 'no'. Let **no** be the sentence $\bigwedge\{(\neg\Box_a p_a \wedge \neg\Box_a\neg p_a) \mid a \in \mathcal{A}\}$, which is the sentence that expresses that none of the children knows that she is dirty (i.e. the information expressed by all children answering 'no' at the same time.)

- Finally, we let for each $\mathcal{B} \subseteq \mathcal{A}$, $\mathbf{dirty}(\mathcal{B})$ abbreviate $\bigwedge_{b\in\mathcal{B}} p_b \wedge \bigwedge_{b\notin\mathcal{B}} \neg p_b$. This sentence expresses that all and only the children in $\mathcal{B}$ have dirty foreheads.

We can now express in DES that if exactly m children are dirty and it is commonly known between all children that they can see each other, then it holds that after a common update with the father's statement that at least one of the children is dirty, and commonly updating $m-1$ times with the fact that all children answer 'no', the resulting state is a situation in which all dirty children know that they are dirty. Formally expressed, this boils down to the following statement:

Proposition 4.1 Let $\mathcal{B}$ be a set containing exactly m children ($m \geq 1$), and let $[\phi]^m$ stand for a sequence of $[\phi]\ldots[\phi]$ of m updates with ϕ. Then it holds for all $a \in \mathcal{B}$: $\mathbf{dirty}(\mathcal{B}), \mathbf{vision}, C_{\mathcal{A}}\mathbf{vision} \models [\mathbf{father}]_{\mathcal{A}}[\mathbf{no}]^{m-1}_{\mathcal{A}}\Box_a p_a$

proof: We will provide a syntactical proof of this statement. Although we do not have an axiomatization for the language containing the common knowledge operators $C_{\mathcal{B}}$, it is easy to see that that following axiom is sound, if $b \in \mathcal{B}$:

$$\vdash C_{\mathcal{B}}\phi \to \Box_b(\phi \wedge C_{\mathcal{B}}\phi)$$

We will make use of this axiom in the proof.

The proof of the proposition is by induction on the number of dirty children m. Assume first that only one child, say a, is dirty. In classical modal logic, it holds that if a is the only dirty child, and a can see all other children, then a knows that if at least one child is dirty, it must be herself:

$$\mathbf{dirty}(\{a\}), \mathbf{vision} \vdash \Box_a(\mathbf{father} \to p_a)$$

We use axiom 5 to conclude that (omitting the subscript $\mathcal{A}$ from the update operators for legibility):

$$\mathbf{dirty}(\{a\}), \mathbf{vision}, C_{\mathcal{A}}\mathbf{vision} \vdash \Box_a(\mathbf{father} \rightarrow [\mathbf{father}]p_a)$$

whence, by axiom 6

$$\mathbf{dirty}(\{a\}), \mathbf{vision}, C_{\mathcal{A}}\mathbf{vision} \vdash [\mathbf{father}]\Box_a p_a$$

For the induction step, let $\mathcal{B}$ be a set of $m+1$ children, and $a, b \in \mathcal{B}$. It holds by induction hypothesis and the necessitation rule that if $\mathcal{B}'$ has m elements, then

$$\vdash \Box_a((\mathbf{dirty}(\mathcal{B}') \wedge \mathbf{vision} \wedge C_{\mathcal{A}}\mathbf{vision}) \rightarrow [\mathbf{father}][\mathbf{no}]^{m-1}\Box_b p_b)$$

Since $C_{\mathcal{A}}\mathbf{vision} \vdash \Box_a(\mathbf{vision} \wedge C_{\mathcal{A}}\mathbf{vision})$ and $\mathbf{dirty}(\mathcal{B}), \mathbf{vision} \vdash \Box_a(\neg p_a \rightarrow \mathbf{dirty}(\mathcal{B}/\{a\}))$, it follows that

$$\mathbf{dirty}(\mathcal{B}), \mathbf{vision}, C_{\mathcal{A}}\mathbf{vision} \vdash \Box_a((\neg p_a \wedge \mathbf{father}) \rightarrow [\mathbf{father}][\mathbf{no}]^{m-1}\Box_b p_b)$$

from which it follows, using axioms 3 and 5, and the fact that $\vdash \mathbf{no} \rightarrow \neg\Box_b p_b$, that

$$\mathbf{dirty}(\mathcal{B}), \mathbf{vision}, C_{\mathcal{A}}\mathbf{vision} \vdash \Box_a(\mathbf{father} \rightarrow [\mathbf{father}][\mathbf{no}]^{m-1}(\mathbf{no} \rightarrow p_a))$$

By axiom 6, then

$$\mathbf{dirty}(\mathcal{B}), \mathbf{vision}, C_{\mathcal{A}}\mathbf{vision} \vdash [\mathbf{father}]\Box_a[\mathbf{no}]^{m-1}(\mathbf{no} \rightarrow p_a)$$

and using the lemma below, finally,

$$\mathbf{dirty}(\mathcal{B}), \mathbf{vision}, C_{\mathcal{A}}\mathbf{vision} \vdash [\mathbf{father}][\mathbf{no}]^m\Box_a p_a$$

.

Lemma 4.2 For each m: $\Box_a[\mathbf{no}]^m(\mathbf{no} \rightarrow p_a) \vdash [\mathbf{no}]^{m+1}\Box_a p_a$

This is proven by induction on m. If $m = 0$, then $\Box_a(\mathbf{no} \rightarrow p_a)$ is equivalent by axiom 5 to $\Box_a(\mathbf{no} \rightarrow [\mathbf{no}]p_a)$, which is, by axiom 6, equivalent to $[\mathbf{no}]\Box_a p_a$.

For the induction step, assume that:

$$\Box_a[\mathbf{no}]^{m+1}(\mathbf{no} \rightarrow p_a)$$

This implies that

$$\Box_a(\mathbf{no} \rightarrow [\mathbf{no}]^{m+1}(\mathbf{no} \rightarrow p_a))$$

From which it follows by axiom 6 that

$$[\mathbf{no}]\Box_a[\mathbf{no}]^m(\mathbf{no} \rightarrow p_a)$$

whence, by induction hypothesis, that

$$[\mathbf{no}][\mathbf{no}]^{m+1}\Box_a p_a$$

This completes the proof of proposition 4.1.[6] □

Discussion

The puzzle of the dirty children and related puzzles have been discussed relatively extensively in the literature, and several formalizations have been given. Our analysis adds to earlier approaches in an essential way, we believe.

First of all, we have rephrased the informal description of the puzzle in the object language of an independently motivated logic. Something like that has not been done before: all earlier formalizations of the puzzle that we know of consist of a more or less *ad hoc* model of the information and information change involved in the puzzle. That means that each variant of the puzzle that differs from the present one calls for a new analysis and the construction of a new model. The relatively straightforward way in which the puzzle can be formalized in DES suggests that similar problems may be formulated in the same way.

Secondly, the fact that our formalization of the puzzle gives results similar to Barwise's semi-formal results, shows that the paradoxical flavor of the puzzle does not stem from a logical mistake. This suggests strongly that the discrepancy between the ideal situation described in the puzzle and a 'real life' situation should not be explained as a difference in principles of logic, but as a result of the complexity of the reasoning involved in the puzzle and the way it depends on the strong trust they should have in the each other's reasoning capabilities.

Thirdly, the formalization given above makes the role that the father's announcement and the children's answers play quite explicit. For example, one of the 'paradoxical' aspects of the puzzle is that the father's statement seems superfluous at first sight if there two or more dirty children present. In such a situation, each of the children already knows that one of the children is dirty (since everyone can see a dirty child). The formal correlate of this fact is a theorem: $p_a \wedge p_b, \mathbf{vision} \vdash \Box_c \mathbf{father}$. The point of the father's statement lies in the assumption that his announcement makes it common knowledge that at least on child is dirty, which was not the case: $p_a \wedge p_b, \mathbf{vision}, C_{\mathcal{A}}\mathbf{vision} \not\vdash C_{\mathcal{A}}\mathbf{father}$. This observation is not new, but our analysis adds to earlier ones in that it is now possible to formulate such facts in the object language.

Another puzzling aspect of the puzzle that is highlighted in our analysis is the fact that the children keep on saying 'no' until 'suddenly' some children

[6] We have not proven that the children do not know that they are dirty *before* they have answered the question $m-1$ times. To show that, one needs an extra assumption that in the initial possibility, none of the children knows whether she is dirty of not, and that this fact is common knowledge. A proof can then be given along the lines of the proof given here.

answer yes, suggests that each answer supplies new information, although, 'syntactically', the children say the same thing each time. This is directly reflected in our semantics: an update with **no** changes the possibility in a certain fixed way, resulting in a new possibility in which another update with **no** may change the possibility again.

This is an example of the failure in DES of the following principle of *Success*:

$$\text{if } b \in \mathcal{B} \text{ then } \models [\phi]_{\mathcal{B}} \Box_b \phi$$

which states that after a group update with a sentence ϕ, each member in the group knows that ϕ. This is not a property of updates in general, and the example of **no** suggests that this is right.[7]

5 Epistemic propositional dynamic logic

In this section, we use the ideas of dynamic epistemic logic for developing a logic, which we call EPDL, for 'epistemic propositional dynamic logic'. Besides update actions this logic will also have send actions and test actions. The main thrust of this move is very much in the line of our discussion in the previous section: by incorporating more notions, such as send actions, that are crucial for understanding communication, into the logical object language, it becomes possible to formalize processes that otherwise would remain part of the metalanguage.

The language of EPDL may be used to specify or describe the behavior of a group of communicating agents, in a very general sense. The phenomena modeled might be human agents speaking to each other, the behavior of processors in a distributed network, or the behavior of a knowledge base and a human agent querying it.

The idea that extensions or variations of epistemic logic may be used to describe such kind of applications is not new. In computer science, the work of Fagin et al. (1995) and Shoham (1993) on agent oriented programming are prime examples. Another example is the work of McCarty (1990). See also de Rijke (1993).

The basic idea here is to treat the update modalities of the previous sections as programs, and extend the language with certain program operators familiar from PDL (cf. Pratt (1976), Goldblatt (1987)). We consider a language in which there are three kinds of basic programs. Firstly, there are update programs of the form $U(\mathcal{B}, \phi)$, that have the effect that ϕ becomes common knowledge in the group of agents $\mathcal{B}$. Secondly, a program $S(a, \mathcal{B}, \phi)$ will stand for the action of agent a sending the message ϕ to all agents in $\mathcal{B}$. Thirdly, the local tests $?(a, \phi)$ stand for the action of agent a testing whether she knows that ϕ. In addition

[7] For this particular example, we even have the surprising fact that there are possibilities w such that $w \models [\mathbf{no}]_{\mathcal{A}} \Box_b \neg \mathbf{no}$.

to these basic programs, we add the program operators of PDL: composition, union and iteration.

Definition 5.1 (Epistemic Propositional Dynamic Logic)
Given an atomic vocabulary $\mathcal{P}$ and a set of actors $\mathcal{A}$, we define a set of assertions Φ and a set of programs Π by simultaneous induction:

- $\Phi \quad \phi ::= p \mid \neg\phi \mid \phi_1 \wedge \phi_2 \mid \Box_{\mathcal{B}}\phi \mid [\pi]\phi$
- $\Pi \quad \pi ::= U(\mathcal{B},\phi) \mid S(a,\mathcal{B},\phi) \mid ?(a,\phi) \mid \pi_1;\pi_2 \mid \pi_1 \cup \pi_2 \mid \pi^*$
  where $a \in \mathcal{A}$, $\mathcal{B} \subseteq \mathcal{A}$. $\Box$

We will interpret programs as relations on possibilities. Since we have included union in the language, which is interpreted as choice in our semantics, programs will be non-deterministic in general: running a particular program may lead to several different outcomes. That means that, in contrast with the updates described in the previous sections, not all programs will have functional interpretations.

Definition 5.2 (Semantics of EPDL)
An interpretation for EPDL is a function I that assigns to each triple consisting of an actor a, a set of actors $\mathcal{B}$ and a formula ϕ a binary relation between possibilities. Relative to such a interpretation we define the truth conditions of assertions and the interpretation of programs inductively by:

truth conditions

$$\begin{array}{rcl}
I,w \models p & \text{iff} & w(p) = 1 \\
I,w \models \phi \wedge \psi & \text{iff} & I,w \models \phi \text{ and } I,w \models \psi \\
I,w \models \neg\phi & \text{iff} & I,w \not\models \phi \\
I,w \models \Box_a\phi & \text{iff} & \forall v \in w(a) : I,v \models \phi \\
I,w \models [\pi]\phi & \text{iff} & \text{for all } v \text{ with } w[\![\pi]\!]_I v,\ I,v \models \phi
\end{array}$$

program interpretations

$$\begin{array}{rcl}
w[\![S(a,\mathcal{B},\phi)]\!]_I v & \text{iff} & (w,v) \in I(a,\mathcal{B},\phi) \\
w[\![?(a,\phi)]\!]_I v & \text{iff} & w = v \text{ and } I,w \models \Box_a\phi \\
w[\![U(\mathcal{B},\phi)]\!]_I v & \text{iff} & w[\mathcal{B}]v \text{ and for all } b \in \mathcal{B}, \\
& & v(b) = \{w'[\![U(\mathcal{B},\phi)]\!]_I \mid w' \in w(b) \text{ and } I,w' \models \phi\} \\
w[\![\pi_1 \cup \pi_2]\!]_I v & \text{iff} & w[\![\pi_1]\!]_I v \text{ or } w[\![\pi_2]\!]_I v \\
w[\![\pi_1;\pi_2]\!]_I v & \text{iff} & \exists u : w[\![\pi_1]\!]_I u[\![\pi_2]\!]_I v \\
w[\![\pi^*]\!]_I v & \text{iff} & w[\![\pi]\!]_I^* v
\end{array}$$

The interpretation function I plays a role only in the clause for the send actions $S(a, \mathcal{B}, \phi)$. The basic idea is that $I(a, \mathcal{B}, \phi)$ describes the effects of the sending action. Of course, the real interest in sending a message is to exchange information. Thus the real burden is to relate send actions to epistemic effects. This can be achieved in EPDL by formulating extra constraints that relate send actions to update actions and to the information of the actors. Typically, these extra constraints will reflect certain properties of the communication channel, or certain pragmatic rules that the actors follow. For example, the axiom

$$\langle S(a, \mathcal{B}, \phi)\rangle \top \rightarrow \Box_a \phi$$

expresses a sincerity condition for actor a (here $\langle \pi \rangle \phi$ is the existential dual of $[\pi]\phi$): actor a can only send the message ϕ if he has the information that ϕ.

As an example of the application of EPDL to communicative situations we discuss a simple example which is called 'the bit-transmission problem' in Fagin et al. (1995, pp. 107ff.). We consider two agents, a sender s and a receiver r. The sender has a certain piece of information (for example that the value of register x is either 0 or 1) which she wants to communicate to the receiver. We let the proposition p represent the information that x has value 1 (and $\neg p$ that the value is 0). We assume the communication line may be faulty; for simplicity's sake we assume that messages either arrive immediately or are lost forever. Since s cannot be sure that her message has arrived, she will continue sending the message p until she has received an acknowledgment from the receiver that he has gotten her message.

Another way of describing the behavior of s and r is as follows. s will send the message p to r until she knows that r knows that p. As soon as r knows that p, he will send a message 'I know that p' to s. Such descriptions that make use of concept such as 'knowledge' are descriptions that we can fairly straightforwardly translate into our language.

For capturing the behavior of this system in EPDL we first of all need to express what the effect of sending a message is. Under the assumptions we have made, that messages either arrive immediately, or are irredeemably lost, the effect of sending a message ϕ to r can be described by the program $U(r, \phi) \cup ?(r, \top)$: either r's information state is updated with ϕ, or nothing at all happens (the test $?(r, \top)$ will always succeed). The corresponding interpretation function I interprets send actions as follows:

$$I(S(a, \mathcal{B}, \phi) = [\![U(\mathcal{B}, \phi)]\!] \cup Id$$

where Id is the identity relation over possibilities. This interpretation of send actions corresponds to the syntactical characterization:

$$[S(s, r, \phi)]\psi \leftrightarrow [(U(r, \phi) \cup ?(r, \top)]\psi$$

The action of s sending r the value of the bit p is can be described by the program

$(\pi_s{:})\ \ (?(s,p); S(s,r,p)) \cup ((s,\neg p); S(s,r,\neg p)).$

While the receiver can be described as performing the following program:

$(\pi_r{:})\ \ (?(r,p); S(r,s,\Box_r p)) \cup (?(r,\neg p); S(r,s,\Box_r \neg p)).$

We can now formulate statements about such programs in the object language. For example, the statement "If the value of x is 1, then the receiver will eventually (that is, after repeating both programs again and again) know that the x is 1" may be represented by: $\Box_s p \rightarrow \langle(\pi_s;\pi_r)^*\rangle\Box_r p$. In fact, this sentence turns out to be valid under the interpretation I given above.

Properties of programs

We may consider how EPDL programs behave with respect to the properties of situations introduced in definition 3.6. In particular, we may ask which properties are preserved under updating with certain programs.

Definition 5.3 Let S be a class of possibilities, I an interpretation. A program π is persistent over a class S of possibilities under the interpretation I iff $w \in S$ and $w[\![\pi]\!]_I v$ imply $v \in S$. $\Box$

The following result claims that if I interprets all send actions as actions that preserve positive or negative introspection, then all programs will.

Proposition 5.4 If each send action $S(a,\mathcal{B},\phi)$ is persistent over $\mathcal{P}$ ($\mathcal{N}$, $\mathcal{P}\cap\mathcal{N}$ respectively) under I, then each program π is persistent over $\mathcal{P}$ ($\mathcal{N}$, $\mathcal{P}\cap\mathcal{N}$). $\Box$

In general, programs are not persistent over the class of truthful situations, nor over the class of consistent situations. One example is the program π_s above. In a situation where S knows that r does not know the value of the bit, the program π_s may result in a possibility in which r knows the value, but s still believes that r does not know it.

Knowledge programs

As observed by Fagin et al. (1995), one way of looking at a problem like the bit transmission problem is in terms of so called *knowledge programs*. That is, we can model the actors as executing a certain set of instructions of the form $\phi \Rightarrow \alpha$, which are read as "if ϕ then do α", where ϕ is a formula of epistemic logic, and α is some action. Formally, they define a knowledge program for an actor a as a set of these instructions $\mathsf{Pg} = \{\phi_1 \Rightarrow \alpha_1, \ldots, \phi_n \Rightarrow \alpha_n\}$. These programs are interpreted indeterministically by requiring that a performs one of the actions α_i for which the test ϕ_i succeeds.

The models they consider consist of 'local states' of agents standing in a certain relation to each other. These 'local states' are meant to correspond in a relatively direct way to states that the agents may actually be in. Actions are interpreted as operations on such representations, while sentences of modal logic are interpreted in a possible worlds model that is derived from this representation. In addition to this 'two-level' architecture, the model contains an explicit representation of time, and a simple logic of time is added to the language.

It is clear that the in EPDL there are programs similar to these knowledge programs, i.e. programs that make an action conditional on the epistemic state of the actor. We will make a few remarks about how our framework compares with the approach adopted by Fagin et al. One of the most salient differences is that the former has an ontology that is much richer. Although this means that the model allows for distinctions that cannot be drawn in our model and that the behavior of a system can be described in much more detail than in our approach, it also implies, as the authors themselves note, that it is often unclear what part of the behavior of a system should be modeled by what part of the semantics.

In the work of Fagin et al., information change on the level of Kripke structures is a notion that is derived from change in the underlying model. By contrast, we have given an explicit semantics for the notion of epistemic update on this level, thereby providing a semantics in which it is much more clear what is going on. Moreover, this means that we are not restricted to using only S5 models, which in the architecture of the system of Fagin et al. seems unavoidable. This is interesting, because it makes it possible to describe situations in which agents are misinformed about either their environment or about the information of the other agents.

6 Conclusions

In this paper, we have combined techniques from epistemic and dynamic logic to arrive at a logic for describing multi-agent information change. The key concept of dynamic semantics is that the meaning of an assertion is the way in which the assertion changes the information of the hearer. Thus a dynamic epistemic semantics consist in a explicit formal definition of the information change potential of a sentence. We used these ideas to arrive at the system of Dynamic Epistemic Semantics, which is semantics for a language describing information change in a multi-agent setting. This semantics proved useful for analyzing the Muddy Children paradox, and also for giving a semantics for knowledge programs, since it enabled us to model knowledge change by giving an explicit semantics to the triggers of the information change (the latter being the assertions made, or the messages sent). We feel that this is an important extension, since standard approaches to for example the Muddy Children (e.g. Fagin et al. 1995) generally use static epistemic logics like S5 to describe the situation before and after a certain epistemic event, leaving the transition between 'before' and 'after' to considerations in the meta-language.[8] In contrast, in dynamic epistemic logic, epistemic actions like updates are first class citizens of the object language of DES. For one thing, this opens the possibility of making artificial agents a bit more intelligent, by giving them an axiomatics for DEL as their tool for reasoning about knowledge change.

Authors adress:
Department of Philosophy
University of Amsterdam
Nieuwe Doelenstraat 15
1012 CP Amsterdam
The Netherlands
gerbrand@illc.uva.nl
groenev@illc.uva.nl

References

[1] Peter Aczel. *Non-well-founded Sets.* CSLI Lecture Notes, Stanford, 1988.

[2] Jon Barwise. Scenes and other situations. *The Journal of Philosophy*, 78(1):369–397, 1981.

[3] Maarten de Rijke. A system of dynamic modal logic. In *Extending Modal Logic*, PhD thesis, ILLC dissertation series 1993-4, University of Amsterdam, 1993.

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

[5] R. Fagin, J.Y. Halpern, and M.Y. Vardi. A modeltheoretic analysis of knowledge. *Journal of the Association for Computing Machinery*, 39(2):382–428, 1991.

[6] Peter Gärdenfors. *Knowledge in Flux.* The MIT Press, Cambridge (Mass.), 1988.

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

[8] In this respect dynamic epistemic logic is similar to belief Revision Theory (Gärdenfors, 1988) and the work on updates of databases (Katsuno and Mendelzon, 1992). Also see the modal semantics for belief revision of Segerberg (1995). Limitations of space do not allow a comparison with these approaches, although we would like to stress that our focus on higher-order information is a major distinction.

[8] Willem Groeneveld. *Logical Investigations into Dynamic Semantics.* PhD thesis, ILLC dissertation series 1995-18, University of Amsterdam, 1995.

[9] Jan Jaspars. *Calculi for Constructive Communication. A Study of the Dynamics of Partial States.* PhD thesis, ITK, Katholieke Universiteit Brabant, 1994.

[10] H. Katsuno and A. O. Mendelzon. On the difference between updating a knowledge base and revising it. In Peter Gärdenfors, editor, *Belief Revision*, pages 183–203. Cambridge University Press, 1992.

[11] John McCarthy. Formalization of two puzzles involving knowedge. In Vladimir Lifschitz, editor, *Formalizing Common Sense: Papers by John McCarthy*, pages 1–61. Ablex, 1990.

[12] John McCarthy. *Formalizing Common Sense: Papers by John McCarthy.* Ablex, 1990.

[13] Yoram Moses, Danny Dolev, and Joseph Y. Halpern. Cheating husbands and other stories: A case study of knowledge, action, and communication. *Distributed Computing*, 1:167–176, 1986.

[14] V.R. Pratt. Semantical considerations on floyd-hoare logic. *Proc. 17th IEEE Symp. on Foundations of Computer science*, pages 109–121, 1976.

[15] Krister Segerberg. Belief revision from the point of view of doxastic logic. *Bulletin of the IGPL*, 3(4):535–553, 1995.

[16] Y. Shoham. Agent oriented programming. *Artificial Intelligence*, 60(1):51–92, 1993.

[17] Peter van Emde Boas, Jeroen Groenendijk, and Martin Stokhof. The conway paradox: Its solution in an epistemic framework. *Proceedings of the third Amsterdam Montague Symposion*, pages 159–182, 1980.

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