

University of Groningen

Epistemic actions and minimal models

Renardel de Lavalette, Gerard R.; Ditmarsch, Hans van

Published in:
 Institut de Recherche en Informatique de Toulouse IRIT (2002)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
 Publisher's PDF, also known as Version of record

Publication date:
 2002

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Renardel de Lavalette, G. R., & Ditmarsch, H. V. (2002). Epistemic actions and minimal models. In *Institut de Recherche en Informatique de Toulouse IRIT (2002)* (pp. 77-90). University of Groningen, Johann Bernoulli Institute for Mathematics and Computer Science.

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Epistemic actions and minimal models

GERARD R. RENARDEL DE LAVALETTE¹

HANS VAN DITMARSCH²

1 Introduction

This paper is about the dynamics of epistemic models, i.e. multimodal S5 models. We investigate the effect of certain epistemic actions on such models, with special interest for minimality of the resulting models and the stream of information between groups of agents. Our choice of models and actions is inspired by epistemic states and moves that occur in knowledge games like Cluedo. We focus on intrinsic models $M = (W, R, V)$, where worlds $w, v \in W$ are structured objects, carrying enough information to define $(w, v) \in R$ and V_w in terms of w and v .

Our main result is the reduction of epistemic models, resulting from epistemic actions, to *minimal* models (with respect to bisimulation). This proceeds in three steps: the first step corresponds with abstraction from the order of actions, the second step with downward transfer between groups of agents, and the third step with upward knowledge transfer between groups of agents.

1.1 Motivation

The models and actions considered in this paper are inspired by knowledge games. These are games where the players do not have full knowledge of the state of the game (e.g. the distribution of cards), and strive to gain specific information about the game state. A good example is Cluedo, where cards are distributed among the players, while three cards remain face down on the table; the players have

¹University of Groningen, Department of Computing Science, P.O. Box 800, 9700 AV Groningen, The Netherlands; grl@cs.rug.nl

²University of Otago, Department of Computer Science, P.O. Box 56, Dunedin, New Zealand; hans@cs.otago.ac.nz

to determine the identity of the cards on the table by asking and answering questions about the cards they hold. Other examples of knowledge games are Mastermind (find out the combination of colored pawns chosen by your opponent) and Happy Families (try to collect ensembles of four cards). We refer to [4] for an overview of epistemic logic and applications, and to [11, 12] for more information on Cluedo and other knowledge games. Some preliminary remarks concerning the subject matter of this paper appeared in [9].

Our initial motivation was the investigation of Cluedo game states. This shifted to a more general interest in the dynamics of epistemic states, and we see this paper as a report on observations and experiments in an epistemic setting involving the construction devised by Baltag (see [2, 1]). The focus on intrinsic and minimal models reflects a preoccupation with concise and informative representations of epistemic models. Related model representations are studied in e.g. [5] (internal semantics), [7, 6] (non-wellfounded semantics) and [8] (modal structures). Minimal models are often taken for granted; however, executing actions interferes with this, because relevant distinctions may become superfluous after an action. We illustrate this with an example.

EXAMPLE 1.1. Consider two agents a and b that do not know the truth about an atom p . First, b suspects a to have learnt the truth about p (i.e., a learns p or a learns $\neg p$ or nothing happens, and all of this is commonly known). Then, a and b are told that p . For the last action, agent b has to update both the information state where a already knew p and the information state where a didn't know p yet. This results in two indistinguishable states of information, hence a non-minimal model. We give details in Figure 1.1 and Example 1.2,1.6.

2 Models and actions

As usual (see e.g. [3]), a Kripke model for a multimodal logic with agents $a \in A$ and atomic propositions $p \in P$ (A and P nonempty) is a structure $M = \langle W, R, V \rangle$, where $W \neq \emptyset$ is the collection of *worlds*, $R = \{R_a \subseteq W \times W \mid a \in A\}$ is a collection of *accessibility relations* on W , and $V = \{V_w : P \rightarrow \{0, 1\} \mid w \in W\}$ is a collection of *valuations*

in the worlds $w \in W$. We call M an *epistemic* model (or S5 model) whenever the R_a are equivalence relations. All models considered are finite.

Our modal language is

$$\varphi ::= p \mid \neg\varphi \mid \varphi \wedge \psi \mid \Box_a \varphi \mid \Box_B^* \varphi$$

where $p \in P$, $a \in A$, $\emptyset \neq B \subseteq A$. The intended meaning of $\Box_B^* \varphi$ is: group B has common knowledge that φ holds.

The interpretation is defined by

$$\begin{aligned} M, w \models p &= (V_w p = 1) \\ M, w \models \neg\varphi &= M, w \not\models \varphi \\ M, w \models \varphi \wedge \psi &= M, w \models \varphi \ \& \ M, w \models \psi \\ M, w \models \Box_a \varphi &= \forall v \in W ((w, v) \in R_a \Rightarrow M, v \models \varphi) \\ M, w \models \Box_B^* \varphi &= \forall v \in W ((w, v) \in (\bigcup_{a \in B} R_a)^* \Rightarrow M, v \models \varphi) \end{aligned}$$

We observe in passing that a propositional formula (i.e. a formula without modal operators) φ is fully characterised by the collection

$$(1.1) \ T_\varphi =_{\text{def}} \{s \in (P \rightarrow \{0, 1\}) \mid s \models \varphi\}$$

of valuations that make φ true. This will be used later.

A model $M = \langle W, R, V \rangle$ is called *intrinsic* if the worlds $w, w' \in W$ are structured objects that contain the information to define $(w, w') \in R_a$ and $V_w p$ in terms of them. Typically in this paper, worlds have the form $w = \langle s, F_w \rangle$, where s is the valuation V_w of w , and $(v, w) \in R_a$ is defined in terms of F_w . In this case we say that W *represents* M .

As usual, a *bisimulation* between two models $M = \langle W, R, V \rangle$ and $M' = \langle W', R', V' \rangle$ is a nonempty relation $B \subseteq W \times W'$ satisfying, for all w, w' with $w B w'$:

$$\begin{aligned} \forall p \in P (V_w p = V_{w'} p) \\ \forall a \in A \forall v \in W (w R_a v \Rightarrow \exists v' \in W' (v B v' \ \& \ w' R'_a v')) \\ \forall a \in A \forall v' \in W' (w' R'_a v' \Rightarrow \exists v \in W (v B v' \ \& \ w R_a v)) \end{aligned}$$

A model M is *minimal* if it is minimal modulo bisimulation, i.e. the only bisimulation between M and itself is the identity relation¹. Finally we mention the fact that, in the class of finite models, minimal

¹Minimality is a purely structural notion here and not a notion relative to a formula or theory that is modeled, as in approaches to belief revision.

models are exactly the models where every world w has a characteristic formula φ_w :

$$(1.2) \quad M \text{ minimal iff for all } w, w' \in W(M, w' \models \varphi_w \Leftrightarrow w = w').$$

2.1 Actions

In [1], Alexandru Baltag presents a construction to model the effect of an action, which we sketch here, restricting ourselves to epistemic actions which leave the propositional valuation of a world unchanged. An *action structure* is a triple $N = \langle X, Q, \text{pre} \rangle$ with $X \neq \emptyset$ a collection of *action alternatives*, $Q = \{Q_a \mid a \in A\}$ a collection of accessibility relations on X , and pre maps action alternatives $x \in X$ on their precondition. The idea is that the pointed action structure (N, x) represents the action x , but the agents do not know the exact nature of x : for $a \in A$, the $y \in X$ with $(x, y) \in Q_a$ are epistemic alternatives for x . $\text{pre}(x)$ is a precondition: action alternative x can only take place in worlds in which $\text{pre}(x)$ is true. We shall assume that $\text{pre}(x)$ is consistent, i.e. $\text{pre}(x) \not\equiv \perp$. Observe that, in the case of epistemic actions, a propositional precondition is at the same time a postcondition (since valuations are left unchanged).

These intuitions are formalised in the definition of the model $M^N = \langle W', R', V' \rangle$, the *effect* of applying action N in M :

$$\begin{aligned} W' &=_{\text{def}} \{ \langle w, x \rangle \in W \times X \mid M, w \models \text{pre}(x) \} \\ R'_a &=_{\text{def}} \{ (\langle w, x \rangle, \langle w', x' \rangle) \mid (w, w') \in R_a, (x, x') \in Q_a \} \\ V'_{\langle w, x \rangle} &=_{\text{def}} V_w \end{aligned}$$

This construction works for all models, but we apply it here only on S5 models and S5 actions (where the Q_a are equivalence relations). It is not hard to verify that, in that case, the resulting model is S5, too. For a related S5-preserving construction, see [11, 13].

As an illustration, see Figure 1.1 which is based on Example 1.1.

2.2 Propositional and simple actions

Propositional actions are actions where all preconditions are propositional, i.e. contain no modal operators. The order of applying propositional actions is not relevant, in the following sense:

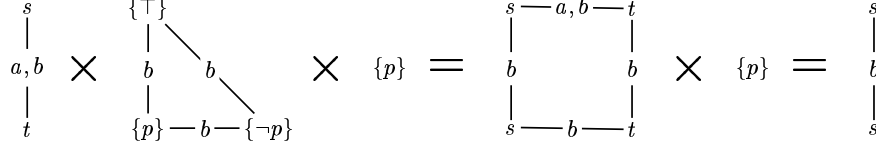


Figure 1.1. The effect of two actions. Worlds are labeled by valuations s, t with $sp = 1, tp = 0$; action alternatives by preconditions. The accessibility relations are equivalence relations and not completely drawn. The last model is bisimilar to the one-world model $\{s\}$. See Example 1.1.

if N_1, N_2 are propositional, then $(M^{N_1})^{N_2}$ and $(M^{N_2})^{N_1}$ are isomorphic, i.e. are bisimilar via a bijective bisimulation.

This follows from the fact that, for propositional φ

$$M^N, \langle w, x \rangle \models \varphi \Leftrightarrow M, w \models \varphi;$$

as a consequence, the collection W' of worlds of $(M^{N_1})^{N_2}$ satisfies

$$W' = \{\langle \langle w, x_1 \rangle, x_2 \rangle \mid M, w \models \text{pre}_1(x_1) \wedge \text{pre}_2(x_2)\}$$

and this is isomorphic with the collection of worlds of $(M^{N_2})^{N_1}$, via the mapping $\langle \langle w, x_1 \rangle, x_2 \rangle \mapsto \langle \langle w, x_2 \rangle, x_1 \rangle$. It is not hard to see that this order-independence only holds for propositional actions, not for epistemic actions in general, let alone for arbitrary actions that may change the propositional valuation in worlds.

Going one step further, we observe that $(M^{N_1})^{N_2}$ and $(M^{N_2})^{N_1}$ are both isomorphic to $M^{N_1 \times N_2}$, where the *product* $N_1 \times N_2 = \langle X, Q, \text{pre} \rangle$ of $N_1 = \langle X_1, Q_1, \text{pre}_1 \rangle$ and $N_2 = \langle X_2, Q_2, \text{pre}_2 \rangle$ is defined by

$$\begin{aligned} X &= X_1 \times X_2 \\ Q_a &= \{\langle \langle x_1, x_2 \rangle, \langle y_1, y_2 \rangle \rangle \mid (x_1, y_1) \in Q_{1,a}, (x_2, y_2) \in Q_{2,a}\} \\ \text{pre}(\langle x_1, x_2 \rangle) &= \text{pre}_1(x_1) \wedge \text{pre}_2(x_2) \end{aligned}$$

This generalises directly to finite sequences of propositional actions. As a consequence, a finite sequence of propositional actions is equivalent to a single propositional action.

Simple actions are epistemic propositional actions where a subgroup $B \subseteq A$ of agents commonly knows which action alternative has taken place; it is publicly known (i.e. the group A of all agents commonly knows) that *some* alternative has taken place. More formally: an action $N = \langle X, Q, \text{pre} \rangle$ is simple iff

1. $\text{pre}(x)$ is propositional for all $x \in X$;
2. for each $a \in A$, Q_a is either maximal (i.e. the universal relation $X \times X$) or minimal (i.e. the identity on X);
3. the *inside group*, i.e. the set of agents a with minimal Q_a , is nonempty.

EXAMPLE 1.2. The first action described in Example 1.1 is a simple action $N = \langle X, Q, \text{pre} \rangle$ with

$$\begin{aligned} X &= \{x_1, x_2, x_3\} \\ Q_a &= \{(x_1, x_1), (x_2, x_2), (x_3, x_3)\}, Q_b = X^2 \\ \text{pre}(x_1) &= \top, \text{pre}(x_2) = p, \text{pre}(x_3) = \neg p \end{aligned}$$

See also Figure 1.1.

EXAMPLE 1.3. Some simple actions, in the context of Cluedo:

- Agent a looks into her (hitherto closed) cards; after this, she (and only she) will know which cards she has, and it will be public knowledge for all agents that she knows. The inside group is $\{a\}$.
- Agent a asks agent b : ‘do you have card k ?’ and b replies ‘no’; now it is public knowledge that b does not have k . The inside group is A .
- Agent a asks agent b : ‘do you have card k or card l ?’ and b shows a one of these cards (k , say) while the other agents do not see which one. Now a and b commonly know that b has k , i.e. $\Box_{\{a,b\}}^*(b \text{ has } k)$, and moreover $\Box_A^*(\Box_{\{a,b\}}^*(b \text{ has } k) \vee \Box_{\{a,b\}}^*(b \text{ has } l))$. The inside group is $\{a, b\}$.

Every simple action $N = \langle X, Q, \text{pre} \rangle$ has a *semantical characterisation* $\langle B, \Sigma \rangle$ where B is the inside group, and $\Sigma = \Sigma_X =_{\text{def}} \{T_{\text{pre}(x)} \mid x \in X\}$ is the collection of sets of valuations corresponding with the (propositional) preconditions of $x \in X$ (see (1.1) for the definition of T_φ).

Observe that the combination of two or more simple actions with identical inside group is again simple. But this is no longer the case with different inside groups. In the sequel, we shall work out more sophisticated semantical characterisations for combinations of simple actions.

We conclude this section by observing that many but not all actions in Cluedo are simple. A counterexample: player a ends her turn without successfully claiming the identity of the cards on the table (and winning the game) (see [11]). This informs the other players that a does not *know* the identity of these cards. Similar actions — learning that some player does not know some proposition from the fact that she does not act in a particular situation — occur in the Muddy Children game (see [4]).

3 Making minimal models

Given a model obtain by simple action execution, we are now going to make it minimal in a three-step contraction based on an intrinsic model representation. First, we have to define such an intrinsic representation. The starting point is some fixed model $M_0 = \langle W_0, R_0, V_0 \rangle$ where the agents have minimal knowledge, i.e. $R_{0,a} = W_0 \times W_0$ for all $a \in A$. (M_0 models the initial state of a knowledge game, e.g. Cluedo: the cards are dealt face down, all players collectively know that nobody knows how the cards are distributed.) Without loss of generality we assume that M_0 is minimal, so different worlds $w \in W_0$ have different valuations $V_{0,w}$. This implies that M_0 is fully characterized by V_0 . The valuations in V_0 represent the game states that are possible, e.g., in the case of Cluedo, the possible distributions of the cards.

Notational convention. From now on, we shall write S for V_0 , and we let s, t range over S .

Let us first see what happens when we apply the simple action

N , characterised by $\langle B, \Sigma \rangle$, to M_0 . We assume that $T \subseteq S$ for all $T \in \Sigma$, i.e. every alternative $T = T_{\text{pre}(x)}$ of N falls within S . Now $M_0^N = \langle W', R', V' \rangle$ with (modulo isomorphism)

$$\begin{aligned} W' &= \{ \langle s, T \rangle \mid s \in T \in \Sigma \} \\ R'_a &= \{ (\langle s, T \rangle, \langle s, T' \rangle) \in W' \times W' \mid a \in B \Rightarrow T = T' \} \\ V'_{\langle s, T \rangle} &= s \end{aligned}$$

So a world $\langle s, T \rangle$ consists of a ‘present state’ s with propositional information, and the alternative $T = T_{\text{pre}(x)}$ that inside group B has learnt. By adding B to the worlds in W' , we obtain the intrinsic representation $\{ \langle s, B, T \rangle \mid s \in T \in \Sigma \}$ of M_0^N .

3.1 A sequence of simple actions

After these preliminaries, we take a number of simple actions N_1, \dots, N_n represented by $\langle B_1, \Sigma_1 \rangle, \dots, \langle B_n, \Sigma_n \rangle$, and apply them to M_0 .

DEFINITION 1.4 (simple model). $M_1 = \langle W_1, R_1, V_1 \rangle$, the result of applying $\langle B_1, \Sigma_1 \rangle, \dots, \langle B_n, \Sigma_n \rangle$ to M_0 , is defined by

$$\begin{aligned} W_1 &= \{ \langle s, \mathbf{B}, \mathbf{T} \rangle \mid s \in \bigcap_{i \leq n} T_i, \forall i \leq n (T_i \in \Sigma_i) \} \\ R_{1,a} &= \{ (\langle s, \mathbf{B}, \mathbf{T} \rangle, \langle s', \mathbf{B}, \mathbf{T}' \rangle) \mid \forall i (a \in B_i \Rightarrow T_i = T'_i) \} \\ V_{1, \langle s, \mathbf{B}, \mathbf{T} \rangle} &= s \end{aligned}$$

Here \mathbf{B}, \mathbf{T} abbreviate (B_1, \dots, B_n) and (T_1, \dots, T_n) , respectively.

Without proof, we state:

PROPOSITION 1.5. M_1 is bisimilar with $M_0^{N_1 \times \dots \times N_n}$.

Observe that M_1 is intrinsic. We may paraphrase the epistemic content of world $w = \langle s, \mathbf{B}, \mathbf{T} \rangle$ as follows: the ‘present state’ is s , and for $i = 1$ to n , group B_i has common knowledge that the present state is some element of T_i .

EXAMPLE 1.6. We go on with Example 1.1 as represented in Figure 1.1. The composition of the two actions is represented by

$$(\{ \{a\}, \{ \{s\}, \{t\}, \{s, t\} \} \}, \{ \{a, b\}, \{ \{s\} \} \})$$

The worlds in the resulting model are $\langle s, (\{a\}, \{a, b\}), (\{s, t\}, \{s\}) \rangle$ and $\langle s, (\{a\}, \{a, b\}), (\{s\}, \{s\}) \rangle$. Note that p is satisfied in both worlds, hence the model is bisimilar with the singleton model where p holds.

3.2 Abstracting from the order of actions

M_1 is an intrinsic representation for M_0 , but it is in general not minimal. To illustrate this, we give a new example:

EXAMPLE 1.7. As in Example 1.1, agents a and b do not know the truth about p , and b suspects a of learning whether p . Then (unlike Example 1.1) a learns whether p . As in Example 1.6 and Figure 1.1, we represent the two valuations of p by s and t . The following sequence of two actions is executed:

$$(\langle \{a\}, \{\{s\}, \{t\}, \{s, t\}\} \rangle, \langle \{a\}, \{\{s\}, \{t\}\} \rangle)$$

The model resulting from this is $M_1 = \langle W_1, R_1, V_1 \rangle$ with

$$\begin{aligned} W_1 = & \{ \langle s, (\{a\}, \{s, t\}), (\{a\}, \{s\}) \rangle, \\ & \langle s, (\{a\}, \{s\}), (\{a\}, \{s\}) \rangle, \\ & \langle t, (\{a\}, \{s, t\}), (\{a\}, \{t\}) \rangle, \\ & \langle t, (\{a\}, \{t\}), (\{a\}, \{t\}) \rangle \} \end{aligned}$$

and with $R_{1,a}$ the identity and $R_{1,b}$ the universal relation. Observe that this is not a minimal model, for the first two worlds are bisimilar (idem for the last two worlds).

The most obvious shortcoming is that the representation of worlds in M_1 depends on the order of actions, while we have seen that the order of simple actions is irrelevant. Another point is: what to do when, for different i, j , $B_i = B_j$? Then B_i has learnt that the present state is both in T_i and in T_j , i.e. in $T_i \cap T_j$. This suggests to combine, for any group B , all T_i with $B_i = B$ and to take their intersection. An orderly way to do this is with functions

$$F : \wp^+(A) \rightarrow \wp^+(S)$$

where $F(B)$ is the intersection of all T_i with $B_i = B$; if no such $i \leq n$ exists, we take the default value S for $F(B)$. The idea is that F corresponds to (the effect of) one of the action alternatives in the composite action $N = N_1 \times \dots \times N_n$: for $B \subseteq A$, the agents in B have common knowledge that the present state is in $F(B)$. We require that F is *consistent*, i.e.

$$\cap F \neq \emptyset, \text{ where } \cap F =_{\text{def}} \bigcap_{B \subseteq A} F(B)$$

The collection of all alternatives F in composite action N is represented by $\Phi = \Phi_N : \wp^+(A) \rightarrow \wp^+(\wp^+(S))$, defined by

$$\Phi(B) = \sqcap \{ \Sigma_i \mid i \leq n, B_i = B \}$$

where \sqcap is defined by

$$\Sigma_1 \sqcap \Sigma_2 =_{\text{def}} \{ T_1 \cap T_2 \mid T_1 \in \Sigma_1, T_2 \in \Sigma_2, T_1 \cap T_2 \neq \emptyset \}$$

and $\sqcap \{ \Sigma_1, \dots, \Sigma_n \} = \Sigma_1 \sqcap \dots \sqcap \Sigma_n$. If, in the definition of $\Phi(B)$, there is no i with $B_i = B$, we take the default value $\{S\}$ for $\sqcap \emptyset$.

So Φ is an order-free representation of the sequence of actions N . We define

$$F \varepsilon \Phi =_{\text{def}} \forall B \subseteq A (FB \in \Phi B) \ \& \ \cap F \neq \emptyset$$

and we shall identify Φ with the collection $\{F \mid F \varepsilon \Phi\}$ of alternatives.

Before defining the new representation of M_1 based on Φ , we introduce some notation. We define $=_B$, ‘equality from B and upward’, by

$$F =_B G =_{\text{def}} \forall C \supseteq B \ FC = GC$$

and write $=_a$ for $=_{\{a\}}$. $F \subseteq_B G$ is defined similarly.

DEFINITION 1.8 (order independent model).

$M_2 = M_2(\Phi) = \langle W_2, R_2, V_2 \rangle$ is defined by

$$\begin{aligned} W_2 &=_{\text{def}} \{ \langle s, F \rangle \mid \forall B \subseteq A \ s \in FB \in \Phi B \} \\ R_2 a &=_{\text{def}} \{ (\langle s, F \rangle, \langle t, G \rangle) \mid F =_a G \} \\ V_2, \langle s, F \rangle &=_{\text{def}} \ s \end{aligned}$$

So a cannot distinguish between the alternatives F and G iff, in all actions involving an inside group B containing a , F and G yield the same information $FA = GA$. We have

PROPOSITION 1.9. M_1 and M_2 are bisimilar, via the bisimulation

$$\langle s, \mathbf{B}, \mathbf{T} \rangle \mapsto \langle s, \lambda B. \cap \{ T_i \mid i \leq n, B_i = B \} \rangle$$

Proof. Straightforward. ■

The order independent model of Example 1.7 is isomorphic to the minimal model $s-b-t$. The domain W_2 is $\{ \langle s, F \rangle, \langle t, G \rangle \}$ with $F(\{a\}) = \{s\}$, $G(\{a\}) = \{t\}$, and $F(B) = G(B) = \{s, t\}$ for $B \neq \{a\}$.

3.3 Information moving downward

In general, the order independent model M_2 is not minimal. The reason is that some values of F may be too large: FB may contain states s, t which can be distinguished by group B , and this is not in line with the intuition behind the representation involving F . This is the case in the order independent representation M_2 of the model of Example 1.1, where $W_2 = \{\langle s, F \rangle, \langle s, G \rangle\}$ with $F(\{b\}) = F(\{a, b\}) = G(\{a\}) = G(\{b\}) = G(\{a, b\}) = \{s\}$ but $F(\{a\}) = \{s, t\}$, although t is no longer an alternative for a .

In this and the next subsection, we will have a closer look at the *stream of information* between groups, and reduce FB via $\overline{FB} \subseteq FB$ to $\tilde{FB} \subseteq \overline{FB}$. We shall show that the representation M_4 based on \tilde{F} is indeed minimal.

The first observation is: if group B learns that the present state is in FB , then this is also learnt by *all subgroups* $C \subseteq B$. So there is information streaming downward, from B to its subgroups. To reflect this, we define the downward closure \overline{F} of F by

$$\overline{FC} =_{\text{def}} \bigcap_{B \supseteq C} FB$$

Observe that \overline{F} is monotonic: if $C \subseteq B$ then $\overline{FC} \subseteq \overline{FB}$, i.e. C (considering less alternatives possible than B) knows more than B .

DEFINITION 1.10 (downward model). $M_3 = M_3(\Phi) = \langle W_3, R_3, V_3 \rangle$ is defined by

$$\begin{aligned} W_3 &=_{\text{def}} \{ \langle s, \overline{F} \rangle \mid F \varepsilon \Phi, s \in \cap F \} \\ R_{3,a} &=_{\text{def}} \{ (\langle s, \overline{F} \rangle, \langle t, \overline{G} \rangle) \mid \overline{F} =_a \overline{G} \} \\ V_{3, \langle s, \overline{F} \rangle} &=_{\text{def}} s \end{aligned}$$

PROPOSITION 1.11. M_2 and M_3 are bisimilar, via the bisimilarity $\langle s, F \rangle \mapsto \langle s, \overline{F} \rangle$.

Proof. Straightforward. ■

EXAMPLE 1.12. The downward model representation of the model of Example 1.1, 1.2 and 1.6 has a singleton domain, containing the

world $\langle s, \overline{F} \rangle$ with $\overline{F}(\{a\}) = \overline{F}(\{b\}) = \overline{F}(\{a, b\}) = \{s\}$. The knowledge acquired by group $\{a, b\}$ in the second action has now moved downward to the individual agents a, b . The model is now minimal.

3.4 Information moving upward

It is tempting to conjecture that M_3 is a minimal model, i.e. different worlds are really different and represent different epistemic alternatives, but that is not the case. We illustrate this with an example.

EXAMPLE 1.13. There are three agents a, b, c and two atoms p and q . Agents a, b, c are in a dark room, a, b wear glasses and c is blindfolded. The glasses are black or transparent. Atom p represents that a wears transparent glasses, atom q represents that b wears transparent glasses. Now the light is turned on, so a and b see which type of glasses they wear themselves, and only agents with transparent glasses can see what the other wears. This can be modeled by two simple actions. First a learns one of $\{p \wedge q\}, \{p \wedge \neg q\}, \{\neg p \wedge q, \neg p \wedge \neg q\}$, shortened by $\{1\}, \{2\}, \{3, 4\}$. Then b learns $\{1\}, \{3\}, \{2, 4\}$. The result of applying these actions is a model with four worlds. Now c suspects that someone else tells a, b that both wear transparent glasses, i.e. that a, b learn $\{1\}$ – both transparent – or $\{1, 2, 3, 4\}$ – nothing happens. The resulting model has five worlds (writing a for $\{a\}$, etc.):

$$\begin{aligned}
 w_0 &= \langle 1, (a, \{1\}), (b, \{1\}), (\{a, b\}, \{1\}) \rangle \\
 w_1 &= \langle 1, (a, \{1\}), (b, \{1\}), (\{a, b\}, \{1, 2, 3, 4\}) \rangle \\
 w_2 &= \langle 2, (a, \{2\}), (b, \{2, 4\}), (\{a, b\}, \{1, 2, 3, 4\}) \rangle \\
 w_3 &= \langle 3, (a, \{3, 4\}), (b, \{3\}), (\{a, b\}, \{1, 2, 3, 4\}) \rangle \\
 w_4 &= \langle 4, (a, \{3, 4\}), (b, \{2, 4\}), (\{a, b\}, \{1, 2, 3, 4\}) \rangle
 \end{aligned}$$

All worlds are indistinguishable for c . The order independent and the downward representation are isomorphic to this model, which is not minimal: w_0 and w_1 are bisimilar.

Let us analyse the subtle process of information streaming upward, from individual agents to groups. Consider $s, t \in \overline{F}\{a, b\}$; then $\{a, b\}$, as a group, has not learnt anything to distinguish s from t , for $s, t \in FB$ for all groups B containing a and b . However, it is possible that the group $\{a, b\}$ can distinguish s from t . This may sound surprising,

but consider the situation that it is publicly known that both a and b have, as individual agents, learnt something to distinguish s from t : then group $\{a, b\}$ has indeed common knowledge to distinguish s from t .

When is it the case that group $\{a, b\}$ cannot distinguish between $s, t \in \overline{F}\{a, b\}$? The answer reads: if there are $s_1 = s, s_2, \dots, s_n = t \in \overline{F}\{a, b\}$ such that it is possible that a cannot distinguish between s_1 and s_2 , b not between s_2 and s_3 , a not between s_3 and s_4 , \dots , and b not between s_{n-1} and s_n . In order to formalise the idea *it is possible that certain agents cannot distinguish between ...*, we define $\tilde{\Phi} : \wp^+(A) \rightarrow \wp(S^2)$ by

$$\tilde{\Phi}B =_{\text{def}} \cup\{(\overline{F}a)^2 \mid a \in B, F \varepsilon \Phi\}$$

Now $(s, t) \in \tilde{\Phi}B$ comes down to: it is possible that some $a \in B$ cannot distinguish between s and t (i.e. there is some alternative $F \varepsilon \Phi$ such that $s, t \in \overline{F}a$). With help of $\tilde{\Phi}$ we define \tilde{F} , the final reduction of F :

$$\tilde{F}B =_{\text{def}} \{t \mid \exists s \in \cap F (s, t) \in (\tilde{\Phi}B \cap (\overline{F}B)^2)^*\}$$

In short notation: $\tilde{F}B = (\cap F)(\tilde{\Phi}B \cap (\overline{F}B)^2)^*$. So $\tilde{F}B$ contains only those states that are reachable within $\overline{F}B$ from some state in $\cap F$ via $\overline{G}a$ -steps, with $G \varepsilon \Phi$ and $a \in B$.

Before we define the final model M_4 and show that it is minimal and bisimilar with M_3 , we present some properties of \tilde{F} . One directly observes

$$\tilde{F} \text{ is monotonic, } \tilde{F}B \subseteq \overline{F}B, \tilde{F}a = \overline{F}a, \tilde{F} = \tilde{\tilde{F}} = \tilde{\tilde{\tilde{F}}} = \tilde{\tilde{\tilde{\tilde{F}}}}.$$

Moreover, the relation $(\tilde{\Phi}B \cap (\overline{F}B)^2)^*$ in the definition of \tilde{F} is an equivalence relation, with $\tilde{F}B$ as one of its equivalence classes. A direct consequence of this is

$$\overline{F}B = \overline{G}B \Rightarrow \tilde{F}B = \tilde{G}B \quad \text{or} \quad \tilde{F}B \cap \tilde{G}B = \emptyset$$

We shall combine two alternatives F, G into a third ($F \triangleleft B \triangleright G$), defined by

$$\begin{aligned} (F \triangleleft B \triangleright G)C &= FC && \text{if } B \subseteq C \\ &= GC && \text{if } B \not\subseteq C \end{aligned}$$

We write $(F \triangleleft a \triangleright G)$ for $(F \triangleleft \{a\} \triangleright G)$. We observe that, obviously, $(F \triangleleft B \triangleright G) =_B F$.

Finally we claim

PROPOSITION 1.14. *If $H = (F \triangleleft a \triangleright G)$, then*

$$\tilde{F} =_{\{a\} \cup B} \tilde{G} \Rightarrow \tilde{F} =_a \tilde{H} =_B \tilde{G}$$

Proof. Assume $\tilde{F} =_{\{a\} \cup B} \tilde{G}$. We have $F =_a H$, and this implies $\tilde{F} =_a \tilde{H}$. To obtain $\tilde{H} =_B \tilde{G}$, we shall show

$$\tilde{H} \subseteq_B \overline{G} \quad \text{and} \quad \tilde{G} \subseteq_B \overline{H}$$

for then $\tilde{H} = \tilde{H} \subseteq_B \overline{G} = \tilde{G} = \tilde{G} \subseteq_B \overline{H} = \tilde{H}$. Now let $B' \supseteq B$, then

$$\begin{aligned} & \tilde{H}B' \\ \subseteq & \quad \{\tilde{H} \text{ monotonic, } \tilde{H} \subseteq \overline{H}\} \\ & \tilde{H}(\{a\} \cup B') \cap \overline{H}B' \\ = & \quad \{\text{definition of } H\} \\ & \tilde{F}(\{a\} \cup B') \cap \overline{H}B' \\ = & \quad \{\tilde{F} =_{\{a\} \cup B} \tilde{G}\} \\ & \tilde{G}(\{a\} \cup B') \cap \overline{H}B' \\ \subseteq & \quad \{\tilde{G} \subseteq \overline{G}\} \\ & \overline{G}(\{a\} \cup B') \cap \overline{H}B' \\ \subseteq & \quad \{\text{definition of } \overline{H}\} \\ & \overline{G}(\{a\} \cup B') \cap \bigcap \{HC \mid C \supseteq B', a \notin C\} \\ = & \quad \{\text{definition of } H\} \\ & \overline{G}(\{a\} \cup B') \cap \bigcap \{GC \mid C \supseteq B', a \notin C\} \\ = & \quad \{\text{definition of } \overline{G}\} \\ & \overline{G}B' \end{aligned}$$

so $\tilde{H} \subseteq_B \overline{G}$. $\tilde{G} \subseteq_B \overline{H}$ is proved likewise. ■

DEFINITION 1.15 (upward model). $M_4 = M_4(\Phi) = \langle W_4, R_4, V_4 \rangle$ is defined by

$$\begin{aligned} W_4 & =_{\text{def}} \{ \langle s, \tilde{F} \rangle \mid F \varepsilon \Phi, s \in \cap F \} \\ R_{4,a} & =_{\text{def}} \{ (\langle s, \tilde{F} \rangle, \langle t, \tilde{G} \rangle) \mid \tilde{F} =_a \tilde{G} \} \\ V_{4, \langle s, \tilde{F} \rangle} & =_{\text{def}} s \end{aligned}$$

The characteristic property of M_4 , when compared with the previous models, is that the transitive closure $R_{4,B} = (\bigcup_{a \in B} R_{4,a})^*$ has a straightforward definition, similar to $R_{4,a}$:

$$(1.3) \quad R_{4,B} = \{(\langle s, \tilde{F} \rangle, \langle t, \tilde{G} \rangle) \mid \tilde{F} =_B \tilde{G}\}$$

The inclusion \subseteq is straightforward, and also holds (in appropriate formulation) for M_2 and M_3 . For \supseteq , Proposition 1.14 is required.

PROPOSITION 1.16. M_3 and M_4 are bisimilar, via $\langle s, \overline{F} \rangle \mapsto \langle s, \tilde{F} \rangle$

Proof. It suffices to show

$$\begin{aligned} \langle s, \overline{F} \rangle R_a \langle t, \overline{G} \rangle &\Rightarrow \langle s, \tilde{F} \rangle R_a \langle t, \tilde{G} \rangle \\ \langle s, \tilde{F} \rangle R_a \langle t, \tilde{G} \rangle &\Rightarrow \exists H (\tilde{G} = \tilde{H} \ \& \ t \in \cap H \ \& \ \langle s, \overline{F} \rangle R_a \langle t, \overline{H} \rangle) \end{aligned}$$

Let $\langle s, \overline{F} \rangle R_a \langle t, \overline{G} \rangle$, so $\overline{F} =_a \overline{G}$. So $\tilde{F}a = \overline{F}a = \overline{G}a = \tilde{G}a$, hence $\tilde{F}B \cap \tilde{G}B \neq \emptyset$ for all $B \ni a$, and with $\overline{F} =_a \overline{G}$ we not get $\tilde{F} =_a \tilde{G}$ and conclude $\langle s, \tilde{F} \rangle R_a \langle t, \tilde{G} \rangle$.

Now the second part. Let $\langle s, \tilde{F} \rangle R_a \langle t, \tilde{G} \rangle$, so $\tilde{F} =_a \tilde{G}$; define $H =_{\text{def}} (F \triangleleft a \triangleright G)$, then $\tilde{F} =_a \tilde{H} = \tilde{G}$ by Proposition 1.14, so it suffices to show that $t \in \cap H$. Now $t \in \overline{G}a = \cap G \cap \overline{G}a = \cap G \cap \overline{F}a \subseteq \cap H$, where $\overline{G}a = \overline{F}a$ follows from $\tilde{F} =_a \tilde{G}$. \blacksquare

PROPOSITION 1.17. M_4 is minimal.

Proof. M_4 is finite, so (as we observed in (1.2)) it suffices to give, for every $\langle s, \tilde{F} \rangle \in W_4$, a characterising formula $\kappa_{\langle s, \tilde{F} \rangle}$ with

$$M_4, \langle t, \tilde{G} \rangle \models \kappa_{\langle s, \tilde{F} \rangle} \Leftrightarrow s = t \ \& \ \tilde{F} = \tilde{G}$$

Put

$$\kappa_{\langle s, \tilde{F} \rangle} =_{\text{def}} \theta_s \wedge \bigwedge_B (\Box_B^* \bigvee_{s' \in \tilde{F}B} \theta_{s'} \wedge \bigwedge_{s' \in \tilde{F}B} \Diamond_B^* \theta_{s'})$$

Now, using (1.3), we have $\langle t, \tilde{G} \rangle \models \kappa_{\langle s, \tilde{F} \rangle} \Leftrightarrow s = t \ \& \ \forall B (\tilde{F}B = \cup \{\cap G' \mid \tilde{G} =_B \tilde{G}'\})$ so it suffices to show

$$\tilde{G}B = \cup \{\cap G' \mid \tilde{G} =_B \tilde{G}'\}.$$

The \supseteq part is easy, so we concentrate on the \subseteq part. Let $t \in \widetilde{G}B$; we want G' with $t \in \widetilde{G}'$ and $\widetilde{G} =_B \widetilde{G}'$. Now let H satisfy $t \in \cap H$; such an H always exists. Put $G' =_{\text{def}} (G \triangleleft B \triangleright H)$ then $t \in \cap H \cap \widetilde{G}B \subseteq \cap H \cap \widetilde{G}'B \subseteq \cap G'$. Now we observe $\overline{G} =_B \overline{G}'$ and $t \in \widetilde{G} \cap \widetilde{G}'$, so $\widetilde{G} =_B \widetilde{G}'$. ■

EXAMPLE 1.18. Applying the construction leading to M_4 to the non-minimal model discussed in Example 1.13, we obtain the expected minimal model with $W_4 = \{w_0, w_2, w_3, w_4\}$, i.e. without w_1 .

4 Concluding remarks

We obtained a minimal and intrinsic representation of models that are the effect of sequences of simple actions. We found this representation via an analysis of the implicit information streams between groups of agents.

Given that finite minimal models have characteristic formulae, as used in the proof of proposition 1.17, our construction may help to provide efficient characterizations of finite models. We still have to investigate the equivalence of our descriptions to those in, e.g., [10].

Our minimization techniques may contribute to faster model checking of epistemic formulas in the resulting minimal models. Repeated application of Baltag's construction leads to exponential growth in the number of worlds, so it may pay off to apply minimization techniques. We have not compared the costs of minimizing with the conceivable gains in speed of model checking.

The type of actions considered here (sequences of simple actions) is rather restricted: we excluded exchange of epistemic information, and dynamic effects on the propositional state of a world (a natural example in game terms: exchange of cards). It will be interesting to investigate the effect of these and other non-propositional actions, where the order of the actions plays a role.

BIBLIOGRAPHY

- [1] A. Baltag. A logic for suspicious players: Epistemic actions and belief updates in games. *Bulletin of Economic Research*, 54(1):1–45, 2002.
- [2] A. Baltag, L.S. Moss, and S. Solecki. The logic of public announcements, common knowledge and private suspicions. Extended abstract appeared in Proc. 7th Conf. Theoretical Aspects of Reasoning about Knowledge (TARK98), p. 43 – 56. To appear in *Annals of Pure and Applied Logic*, 2002.

- [3] P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*. Cambridge University Press, Cambridge, 2001. Cambridge Tracts in Theoretical Computer Science 53.
- [4] Ronald Fagin, Joseph Y. Halpern, Yoram Moses, and Moshe Y. Vardi. *Reasoning about Knowledge*. MIT Press, Cambridge, Mass., 1995.
- [5] Ronald Fagin and Moshe Y. Vardi. An internal semantics for modal logic: preliminary report. In *Proceedings of the 7th annual ACM symposium on Theory of Computing*, pages 305 – 315, New York, 1985. ACM.
- [6] Jelle Gerbrandy. *Bisimulations on Planet Kripke*. PhD thesis, Institute for Logic, Language and Computation, Amsterdam, 1999.
- [7] Jelle Gerbrandy and Willem Groeneveld. Reasoning about Information Change. *Journal of Logic, Language and Information*, 6:147–169, 1997.
- [8] Gerard R. Renardel de Lavalette. Modal structures and changing modalities. Manuscript, 2002, see <http://www.cs.rug.nl/~gr1/pub/mscm.ps>
- [9] Gerard R. Renardel de Lavalette. Memories and Knowledge Games. In Maarten de Rijke, Jelle Gerbrandy, Maarten Marx and Yde Venema, editors, *JFAK. Essays Dedicated to Johan van Benthem on the Occasion of his 50th Birthday*. Vossiuspers AUP/Amsterdam University Press, Amsterdam, 1999.
- [10] J.F.A.K. van Benthem. Dynamic odds and ends. Technical report, ILLC, University of Amsterdam, 1998. Report ML-1998-08.
- [11] H.P. van Ditmarsch. *Knowledge games*. PhD thesis, University of Groningen, 2000. ILLC Dissertation Series DS-2000-06.
- [12] H.P. van Ditmarsch. Knowledge games. *Bulletin of Economic Research*, 53(4):249–273, 2001.
- [13] H.P. van Ditmarsch. The description of game actions in cluedo. In L.A. Petrosian and V.V. Mazalov, editors, *Game Theory and Applications*, volume 8, pages 1–28, Commack, NY, USA, 2002. Nova Science Publishers.