

Department of Computer Science,
University of Otago

UNIVERSITY
of
OTAGO



Te Whare Wānanga o Ōtāgo

Technical Report OUCS-2004-18

Playing cards with Hintikka
An introduction to dynamic epistemic logic

Authors:

Hans van Ditmarsch

Department of Computer Science, University of Otago

Wiebe van der Hoek

Department of Computer Science, University of Liverpool

Barteld Kooi

Department of Philosophy, University of Groningen

Status: to appear in the newsletter *Phinews*



Department of Computer Science,
University of Otago, PO Box 56, Dunedin, Otago, New Zealand

<http://www.cs.otago.ac.nz/research/techreports.html>

Playing cards with Hintikka

An introduction to dynamic epistemic logic*

H.P. van Ditmarsch[†] W. van der Hoek[‡] B.P. Kooi[§]

1 Introduction

Imagine three players Anne, Bill, and Cath, each holding one card from a ‘stack’ of three (known) cards clubs, hearts, and spades, such that they only know their own card but do not know which other card is held by which other player. Assume that the actual deal is that Anne holds clubs, Bill holds hearts and Cath holds spades. Now Anne announces that she does not have hearts. What was known before this announcement, and how does this knowledge change as a result of that action? Before, Cath did not know that Anne holds clubs, but afterwards she knows that Anne holds clubs. This is because Cath can reason as follows: “I have spades, so Anne must have clubs or hearts. If she says that she does not have hearts, she must therefore have clubs.” Bill knows that Cath now knows Anne’s card, even though he does not know himself what Anne’s card is. Both before and after, players know which card they hold in their hands. Note that the *only* change that appears to have taken place is epistemic change, and that no factual change has taken place, such as cards changing hands. How do we describe such an information update in an epistemic setting? We can imagine various other actions that affect the knowledge of the players, for example, the action where Anne *shows* her clubs card to Bill, in such a way that Cath sees that Anne is doing that, but without seeing the actual card. How does that affect the knowledge of the players about each other? After that action, Cath still does not know whether Anne holds clubs or hearts. But Cath now knows that Bill knows Anne’s card.

This contribution is a gentle introduction to so-called dynamic epistemic logics, that can describe how agents change their knowledge and beliefs. We start

*Section 5 is partly based on a chapter of [vDvdHK04], and Section 6 is partly based on a section of [vDK04].

[†]Department of Computer Science, University of Otago, Dunedin, New Zealand, hans@cs.otago.ac.nz

[‡]Department of Computer Science, The University of Liverpool, Liverpool, United Kingdom, wiebe@csc.liv.ac.uk

[§]Department of Philosophy, University of Groningen, Groningen, the Netherlands, barteld@philos.rug.nl

with a very concise introduction in epistemic logic, by the example of one, two and finally three players holding cards; and, mainly for the purpose of motivating the dynamics, we also very summarily introduce the concepts of general and common knowledge. We then pay ample attention to the logic of public announcements, wherein agents change their knowledge as the result of, indeed, public announcements. One crucial topic in that setting is that of unsuccessful updates: formulas that become false when announced. The Moore-sentences that were already extensively discussed at the conception of the area of epistemic logic in [Hin62] give rise to such unsuccessful updates. After that, we present a few examples of more complex epistemic updates. Our closing observations are on recent developments that link the ‘standard’ topic of (theory) belief revision ([AGM85]) to the dynamic epistemic logics introduced here.

2 One agent

We introduce epistemic logic by a simple example, even simpler than the one in the introduction. Suppose there is only one player: Anne.

Anne draws one from a stack of three different cards clubs, hearts, and spades. Suppose she draws the clubs card – but she does not look at her card yet; and that one of the remaining cards is put back into the stack holder, suppose that is the hearts card; and that the remaining card is put (face down) on the table. That must therefore be the spades card! Anne now looks at her card.

What does Anne know? We would like to be able to evaluate system descriptions such as:

- Anne holds the clubs card.
- Anne knows that she holds the clubs card.
- Anne does not know that the hearts card is on the table.
- Anne can imagine that the hearts card is on the table.
- Anne knows that the hearts card or the spades card is in the stack holder.
- Anne knows her own card.
- The card on the table is different from the card held by Anne.
- Anne knows that she holds one card.

Facts about the state of the world are in this case facts about card ownership. We describe such facts by atoms such as $Clubs_a$ standing for ‘the clubs card is held by Anne’, and similarly $Clubs_h$ for ‘the clubs card is in the stack holder,

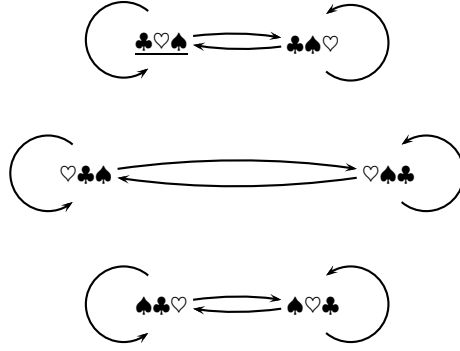


Figure 1: A pointed Kripke model, also known as epistemic state, that represents Anne's knowledge for the card deal where Anne holds clubs, hearts is in the stack holder, and spades is on the table. The actual state is underlined.

and $Clubs_t$ for 'the clubs card is on the table', etc. The standard propositional connectives are \wedge for 'and', \vee for 'or', \neg for 'negation', \rightarrow for 'implication', and \leftrightarrow for 'equivalence'. A formula of the form $K\varphi$ expresses that 'Anne knows that φ ', and a formula of the form $\hat{K}\varphi$ (\hat{K} is the dual of K) expresses that 'Anne can imagine that φ '. The informal descriptions above become

- Anne holds the clubs card: $Clubs_a$
- Anne knows that she holds the clubs card: $KClubs_a$
- Anne does not know that the hearts card is on the table: $\neg KHearts_t$
- Anne can imagine that the hearts card is not on the table: $\hat{K}\neg Hearts_t$
- Anne knows that the hearts card or the spades card are in the stack holder: $K(Hearts_h \vee Spades_h)$
- Anne knows her own card: $KClubs_a \vee KHearts_a \vee KSpades_a$
- The card on the table is not held by Anne: $(Clubs_t \rightarrow \neg Clubs_a) \wedge (Hearts_t \rightarrow \neg Hearts_a) \wedge (Spades_t \rightarrow \neg Spades_a)$
- Anne knows that she holds one card: $K((Clubs_a \rightarrow (\neg Hearts_a \wedge \neg Spades_a)) \wedge (Hearts_a \rightarrow (\neg Clubs_a \wedge \neg Spades_a)) \wedge (Spades_a \rightarrow (\neg Hearts_a \wedge \neg Clubs_a)))$

So far, so good. Now how are we going to interpret those formulas? The operator K can be interpreted as a modal operator, of the 'necessity'- or \Box -type, on structures that are Kripke models. Formally, an epistemic state, or information state, is a pointed relational structure consisting of a set of 'states of the world', a binary relation of 'accessibility' between states, and a factual description of the states – i.e., a valuation of facts on all states. In our example, the states are card deals. The deal where Anne holds the clubs card, the hearts card is in the stack holder and the spades card is on the table, we give the

name $\clubsuit\heartsuit\spadesuit$, etc. By identifying states with deals, we have implicitly specified the evaluation of facts in the state with the name $\clubsuit\heartsuit\spadesuit$. The binary relation of accessibility between states expresses what the player knows about the facts. For example, if deal $\clubsuit\heartsuit\spadesuit$ is actually the case, Anne holds the clubs card, and in that case she can imagine that not $\clubsuit\heartsuit\spadesuit$ but $\clubsuit\clubsuit\heartsuit$ is the case, wherein she also holds the clubs card. We say that state $\clubsuit\clubsuit\heartsuit$ is accessible from state $\clubsuit\heartsuit\spadesuit$ for Anne, or that $(\clubsuit\heartsuit\spadesuit, \clubsuit\clubsuit\heartsuit)$ is in the accessibility relation. Also, she can imagine the actual deal $\clubsuit\heartsuit\spadesuit$ to be the case, so $\clubsuit\heartsuit\spadesuit$ is ‘accessible from itself’: the pair $(\clubsuit\heartsuit\spadesuit, \clubsuit\heartsuit\spadesuit)$ must also be in the accessibility relation.

Continuing in this way, we get access as in Figure 1. This structure can formally be described as a pointed Kripke model $(Hexa_a, \clubsuit\heartsuit\spadesuit)$ where the model $Hexa_a = \langle S, R, V \rangle$ consists of a domain S , accessibility relation R and valuation V such that

$$\begin{aligned} S &= \{\clubsuit\heartsuit\spadesuit, \clubsuit\clubsuit\heartsuit, \heartsuit\clubsuit\spadesuit, \heartsuit\heartsuit\clubsuit, \spadesuit\clubsuit\heartsuit, \spadesuit\heartsuit\clubsuit\} \\ R &= \{(\clubsuit\heartsuit\spadesuit, \clubsuit\heartsuit\spadesuit), (\clubsuit\heartsuit\spadesuit, \clubsuit\clubsuit\heartsuit), (\clubsuit\clubsuit\heartsuit, \clubsuit\clubsuit\heartsuit), \dots\} \\ V(Clubs_a) &= \{\clubsuit\heartsuit\spadesuit, \clubsuit\clubsuit\heartsuit\} \\ V(Hearts_a) &= \{\heartsuit\clubsuit\spadesuit, \heartsuit\heartsuit\clubsuit\} \\ &\dots \end{aligned}$$

The states where a given atom is true, are identified with a subset of the domain: $Clubs_a$ – for ‘Anne holds the clubs card’ – is only true in states $\{\clubsuit\heartsuit\spadesuit, \clubsuit\clubsuit\heartsuit\}$, etc. A standard modal language inductively defined by $\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \psi) \mid \Box\varphi$ can now interpreted on this structure – let’s stick to the familiar \Box for a little while, before we write K for that. The crucial clause in the interpretation of formulas is the one for the modal operator: $M, s \models \Box\varphi$ if and only if for all s' , if $R(s, s')$, then $M, s' \models \varphi$. For $M, s \models \varphi$ read ‘state s of model M satisfies formula φ ’, or ‘ φ is true in state s of model M ’. For example, we can now compute that in the epistemic state $(Hexa_a, \clubsuit\heartsuit\spadesuit)$ it is indeed true that Anne knows that she holds the clubs card:

We have that $Hexa_a, \clubsuit\heartsuit\spadesuit \models \Box Clubs_a$ if and only if [for all states s , if $R(\clubsuit\heartsuit\spadesuit, s)$ then $Hexa_a, s \models Clubs_a$]. The last is implied by $Hexa_a, \clubsuit\heartsuit\spadesuit \models Clubs_a$ and $Hexa_a, \clubsuit\clubsuit\heartsuit \models Clubs_a$, as the only states that are accessible from $\clubsuit\heartsuit\spadesuit$ are $\clubsuit\heartsuit\spadesuit$ itself and $\clubsuit\clubsuit\heartsuit$: we have $R(\clubsuit\heartsuit\spadesuit, \clubsuit\heartsuit\spadesuit)$ and $R(\clubsuit\heartsuit\spadesuit, \clubsuit\clubsuit\heartsuit)$. Finally, $Hexa_a, \clubsuit\heartsuit\spadesuit \models Clubs_a$ because $\clubsuit\heartsuit\spadesuit \in V(Clubs_a) = \{\clubsuit\heartsuit\spadesuit, \clubsuit\clubsuit\heartsuit\}$, and, similarly, $Hexa_a, \clubsuit\clubsuit\heartsuit \models Clubs_a$ because $\clubsuit\clubsuit\heartsuit \in V(Clubs_a) = \{\clubsuit\heartsuit\spadesuit, \clubsuit\clubsuit\heartsuit\}$. Done! From now on, we will always write K for \Box .

It turns out that Anne’s accessibility relation is an equivalence relation. If one assumes certain properties of knowledge, this is always the case. The properties are that ‘what you know is true’, which is formalized by the schema $K\varphi \rightarrow \varphi$; that ‘you are aware of your knowledge’, which is formalized by the schema $K\varphi \rightarrow KK\varphi$, and that ‘you are aware of your ignorance’, which is formalized by the schema $\neg K\varphi \rightarrow K\neg K\varphi$. These properties may be disputed for various reasons, for example, without the requirement that what you know is true, we get a notion of belief instead of knowledge. For now, also for the sake of a simple exposition, we will stick to the properties of knowledge and see where they get us.

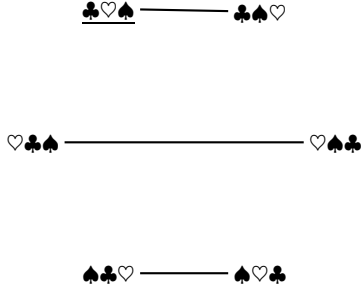


Figure 2: A simpler visualization for the epistemic state where Anne holds clubs, hearts is in the stack holder, and spades is on the table. The actual state is underlined.

Together, they enforce that in epistemic logic the accessibility relation is always an equivalence relation. This is somewhat differently expressed, by saying that what a player / agent cannot distinguish from one another induces a *partition* on the set of states, i.e., a set of equivalence classes that cover the entire domain. For equivalence relations, we write \sim instead of R , and we write this ‘infix’, i.e., we write $\clubsuit\heartsuit\spadesuit \sim \clubsuit\spadesuit\heartsuit$ instead of $R(\clubsuit\heartsuit\spadesuit, \clubsuit\spadesuit\heartsuit)$. In the case of equivalence relations a simpler visualization is sufficient: we only need to link visually the states that are in the same class. If a state is not linked to others, it must be a singleton equivalence class (reflexivity always holds). For $(Hexa_a, \clubsuit\heartsuit\spadesuit)$ we get the visualization in Figure 2.

One might ask: why not restrict ourselves in the model to the two deals $\clubsuit\heartsuit\spadesuit$ and $\clubsuit\spadesuit\heartsuit$ only? The remaining deals are inaccessible anyway from the actual deal! From an agent’s point of view this is arguably right, but from a modeller’s point of view the six-point model is preferable: this model works regardless of the actual deal.

The dual of ‘know’ is ‘can imagine that’: $\hat{K}\varphi := \neg K\neg\varphi$, so that ‘can imagine that’ means ‘not knowing that not’. For example, ‘Anne can imagine that the hearts card is not on the table’ is described by $\hat{K}\neg Hearts_t$ which is true in epistemic state $(Hexa_a, \clubsuit\heartsuit\spadesuit)$, because from deal $\clubsuit\heartsuit\spadesuit$ Anne can access deal $\clubsuit\spadesuit\heartsuit$ for which $\neg Hearts_t$ is true, as the spades card is on the table in that deal. There appears to be no generally accepted notation for ‘can imagine that’. The ‘hat’ in the notation $\hat{K}\varphi$ – the notation we will keep using – is reminiscent of the diamond in $\diamond\varphi$; alternatively one may find for that, in the literature, $M\varphi$ or $k\varphi$.

3 More agents

Much formal dynamics can be presented based on the single-agent situation. For example, the action of Anne picking up the card from the table that has been dealt to her, is a significantly complex epistemic action. But a proper

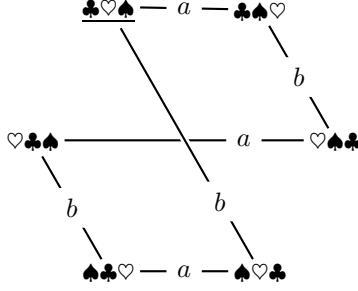


Figure 3: Anne and Bill both draw one from the cards clubs, hearts, and spades. The remaining card is put (face down) on the table. Anne draws the clubs card and Bill draws the hearts card.

and more interesting perspective is that of the multi-agent situation. This is, because players may now have knowledge about each others' knowledge, so that for even a single fact the Kripke models representing that knowledge can become arbitrarily complex. For a start, let's move from one to two players in the three cards situation:

Anne and Bill both draw one from the cards clubs, hearts, and spades. The remaining card is put (face down) on the table. Suppose Anne draws the clubs card and Bill draws the hearts card.

The epistemic operator K with corresponding access \sim , to describe Anne's knowledge, now has to be different from an epistemic operator and corresponding access for Bill. The distinction can easily be made by labelling an operator, and access, with the agent that it is knowledge and access for. If we take a for Anne, and b for Bill, this results in equivalence relations \sim_a and \sim_b and corresponding knowledge operators K_a and K_b . Bill's access on the domain is different from Anne's: whereas Anne cannot tell deals $\clubsuit\heartsuit\spadesuit$ and $\clubsuit\spadesuit\heartsuit$ apart, Bill instead cannot tell deals $\clubsuit\heartsuit\spadesuit$ and $\spadesuit\heartsuit\clubsuit$ apart, etc. The resulting model Hex_{ab} is visualized in Figure 3. We can now describe in the epistemic language that, for example:

- Bill cannot imagine that Anne has the hearts card: $\neg\hat{K}_b\text{Hearts}_a$
- Anne can imagine Bill to imagine that she has the hearts card: $\hat{K}_a\hat{K}_b\text{Hearts}_a$
- Anne knows Bill to imagine that she has the clubs card: $K_a\hat{K}_b\text{Clubs}_a$

The formula $\hat{K}_a\hat{K}_b\text{Hearts}_a$ is true in epistemic state $(Hex_{ab}, \clubsuit\heartsuit\spadesuit)$ – formally, $(Hex_{ab}, \clubsuit\heartsuit\spadesuit) \models \hat{K}_a\hat{K}_b\text{Hearts}_a$. This can be shown as follows. We have that $\clubsuit\heartsuit\spadesuit \sim_a \clubsuit\spadesuit\heartsuit$ and that $\clubsuit\spadesuit\heartsuit \sim_b \heartsuit\spadesuit\clubsuit$. In the last state, we have $(Hex_{ab}, \heartsuit\spadesuit\clubsuit) \models \text{Hearts}_a$. From that and $\clubsuit\spadesuit\heartsuit \sim_b \heartsuit\spadesuit\clubsuit$ follows $(Hex_{ab}, \clubsuit\spadesuit\heartsuit) \models \hat{K}_b\text{Hearts}_a$, and from that and $\clubsuit\heartsuit\spadesuit \sim_a \clubsuit\spadesuit\heartsuit$ follows $(Hex_{ab}, \clubsuit\heartsuit\spadesuit) \models \hat{K}_a\hat{K}_b\text{Hearts}_a$.

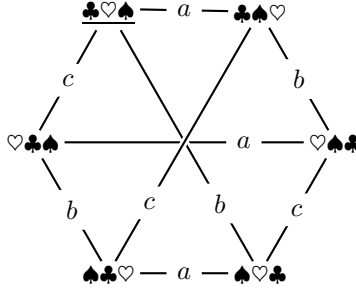


Figure 4: The epistemic state ($Hexa, \clubsuit\heartsuit\spadesuit$) for the card deal where Anne holds clubs, Bill holds hearts, and Cath holds spades.

For three cards and three agents, we get the model $Hexa$ pictured in Figure 4, and we can now describe in the epistemic language that:

- Anne knows that Bill knows that Cath knows her own card: $K_a K_b (K_c Clubs_c \vee K_c Hearts_c \vee K_c Spades_c)$
- Anne has the clubs card, but Anne can imagine that Bill can imagine that Cath knows that Anne does not have the clubs card: $Clubs_a \wedge \hat{K}_a \hat{K}_b K_c \neg Clubs_a$

The structures we will use throughout this presentation can now be introduced formally as follows:

Definition 1 (Epistemic structures)

An *epistemic model* $M = \langle S, \sim, V \rangle$ consists of a *domain* S of (factual) *states* (or ‘worlds’), *accessibility* $\sim : N \rightarrow \mathcal{P}(S \times S)$, and a *valuation* $V : P \rightarrow \mathcal{P}(S)$. For $s \in S$, (M, s) is an *epistemic state*.

For $\sim(n)$ we write \sim_n and for $V(p)$ we write V_p . So, access \sim can be seen as a set of equivalence relations \sim_n , and V as a set of valuations V_p . Relative to a set of agents N and a set of atoms P , the language of multiagent epistemic logic is inductively defined by $\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \psi) \mid K_n\varphi$. We need some further extensions of the language, but all these will be interpreted on the structures presented in Definition 1.

4 Common knowledge

The first extension of the language is with epistemic operators for *groups* of agents, specifically: with common knowledge. As we aim to focus on *dynamic* epistemics in this contribution, and not on dynamic *epistemics*, this will be a lightning quick introduction in those concepts. For more information, see [FHMV95, MvdH95].

In the epistemic state ($Hexa, \clubsuit\heartsuit\spadesuit$) of Figure 4 both Anne and Bill know that the deal of cards is *not* $\spadesuit\clubsuit\heartsuit$: both $K_a\neg(\text{Spades}_a \wedge \text{Clubs}_b \wedge \text{Hearts}_c)$ and $K_b\neg(\text{Spades}_a \wedge \text{Clubs}_b \wedge \text{Hearts}_c)$ are true. If a group of agents all individually know that φ , we say that φ is general knowledge. The modal operator for general knowledge of a group G is E_G . For an arbitrary subset $G \subseteq N$ of the set of agents N , we define $E_G\varphi := \bigwedge_{n \in G} K_n\varphi$. So in this case we have that $E_{ab}\neg(\text{Spades}_a \wedge \text{Clubs}_b \wedge \text{Hearts}_c)$ – par abus de langage we write E_{ab} instead of $E_{\{a,b\}}$. Now even though φ may be generally known, that does not imply that agents know *from each other* that they know φ . For example, $K_bK_a\neg(\text{Spades}_a \wedge \text{Clubs}_b \wedge \text{Hearts}_c)$ is *false* in ($Hexa, \clubsuit\heartsuit\spadesuit$): Bill can imagine Anne to have the spades card instead of clubs. In that case, Anne can imagine that the card deal is $\spadesuit\clubsuit\heartsuit$. So $\hat{K}_a\hat{K}_b(\text{Spades}_a \wedge \text{Clubs}_b \wedge \text{Hearts}_c)$ is true, and therefore $K_bK_a\neg(\text{Spades}_a \wedge \text{Clubs}_b \wedge \text{Hearts}_c)$ is false. For other examples, one can construct formulas that are true to some extent $K_aK_bK_cK_aK_aK_b\varphi$ but no longer if one adds one more operator at the start, e.g., $K_bK_aK_bK_cK_aK_aK_b\varphi$ is false. A formula φ is *common knowledge* for a group G , notation $C_G\varphi$, if it holds for arbitrary long stacks of individual knowledge operators (for individuals in that group). If, for example, $G = \{a, b, c\}$, we get something (involving an enumeration of all finite stacks of knowledge operators) like $C_{abc}\varphi := \varphi \wedge K_a\varphi \wedge K_b\varphi \wedge K_c\varphi \wedge K_aK_a\varphi \wedge K_aK_b\varphi \wedge K_aK_c\varphi \wedge \dots K_aK_aK_a\varphi \dots$. Alternatively, we may see common knowledge as the conjunction of arbitrarily many applications of general knowledge: $C_G\varphi := \varphi \wedge E_G\varphi \wedge E_G E_G\varphi \wedge \dots$. Such infinitary definitions are frowned upon. Therefore common knowledge C_G is added as a primitive operator to the language, whereas general knowledge is typically defined (for a finite set of agents) by the notational abbreviation above. Instead, common knowledge is defined semantically, based on transitive closure relations of access for the individual agents in the group. By way of validities involving common knowledge, that are mentioned at the end of this section, any single conjunct from the right side of the infinitary definition of common knowledge is then entailed, and in this way we avoid having to define it in that infinitary way.

The semantics of common knowledge formulas is: $C_G\varphi$ is true in an epistemic state (M, s) , if φ is true in any state s_m that can be reached by a finite path of linked states $s \sim_{n_1} s_1 \sim_{n_2} s_2 \sim_{n_3} \dots \sim_{n_m} s_m$, with all of $n_1, \dots, n_m \in G$ (and not necessarily all different). Mathematically, ‘reachability by a finite path’ is the same as ‘being in the transitive reflexive closure’. If we define \sim_G as $(\bigcup_{n \in G})^*$ – which is that reflexive transitive closure – then we interpret common knowledge as

$$M, s \models C_G\varphi \text{ if and only if for all } s' : s \sim_G s' \text{ implies } M, s' \models \varphi$$

If all individual accessibility relations are equivalence relations, \sim_G is also an equivalence relation. Common knowledge for the entire group N of agents is called *public knowledge*.

In the model $Hexa$, access for any subgroup of two players, or for all three, is the entire model. For such groups G , $C_G\varphi$ is true in an epistemic state $(Hexa, s')$ iff φ is valid on the model $Hexa$ – a formula is valid on a model M , notation

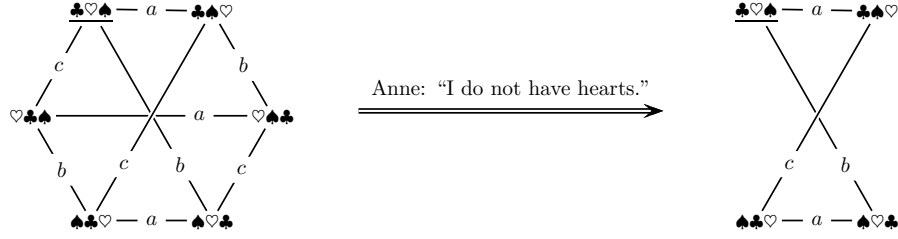


Figure 5: On the left, the epistemic state $(Hexa, \clubsuit\heartsuit\spadesuit)$ for the card deal where Anne hold clubs, Bill holds hearts, and Cath holds spades. The actual deal is underlined. On the right, the effect of Anne saying that she does not have hearts.

$M \models \varphi$, if and only if for all states s in the domain of M : $M, s \models \varphi$. For example, we have that:

- It is public knowledge that Anne knows her card:
 $Hexa \models C_{abc}(K_a Clubs_a \vee K_a Hearts_a \vee K_a Spades_a)$
- Anne and Bill share the same knowledge as Bill and Cath:
 $Hexa \models C_{ab}\varphi \rightarrow C_{bc}\varphi$

Valid principles for common knowledge are $C_G(\varphi \rightarrow \psi) \rightarrow C_G\varphi \rightarrow C_G\psi$ (distribution of C_G over \rightarrow), and $C_G\varphi \rightarrow (\varphi \wedge E_G C_G\varphi)$ (use of C_G), and $C_G(\varphi \rightarrow E_G\varphi) \rightarrow \varphi \rightarrow C_G\varphi$ (induction). Some grasp of group concepts of knowledge is important to understand the effects of public announcements, but we will not pay more attention here to those concepts.

5 Public announcements

We now move on to the dynamics of knowledge. Suppose Anne says that she does not have the hearts card. She then makes public to all three players that all deals where $Hearts_a$ is true can be eliminated from consideration. This results in a restriction of the model $Hexa$ as visualized in Figure 5. The public announcement “I do not have hearts” can be seen as an epistemic ‘program’ with ‘precondition’ $\neg Hearts_a$, that is interpreted as a ‘state transformer’ of the original epistemic state, exactly as a program in *dynamic* modal logic. Given some program π , in dynamic logic $[\pi]\psi$ means that after every execution of π (state transformation induced by π), formula ψ holds. For announcements we want something of the form $[\varphi]\psi$, meaning that after (every) announcement of φ , formula ψ holds.

We *appear* to be moving away slightly from the standard paradigm of modal logic. So far, the accessibility relations were between states in a given model

underlying an epistemic state. But all of a sudden, we are confronted with an accessibility relation between epistemic states as well: “I do not have hearts” induces a(n) (epistemic) state transition such that the pair of epistemic states in Figure 5 is in that relation. The epistemic states take the roles of the points or worlds in a seemingly underspecified domain of ‘all possible epistemic states’. By lifting accessibility between points in the original epistemic state to accessibility between epistemic states, we can get the dynamic and epistemic accessibility relations ‘on the same level’ again, and see this as an ‘ordinary structure’ on which to interpret a perfectly ordinary multimodal logic. A crucial point is that this ‘higher-order structure’ is induced by the initial epistemic state and the actions that can be executed there, and not the other way round. So it’s standard modal logic after all.

Anne’s announcement “I do not have hearts” is a simple epistemic action in various respects. It is public, and therefore not private or even something else. It is truthful, and not merely introspective or even weaker; in that sense it describes change of knowledge only and not change of belief. It is deterministic, i.e. a state transformer; other actions, of which we will see an example, are non-deterministic.

The effect of the public announcement of φ is the restriction of the epistemic state to all worlds where φ holds. So, ‘announce φ ’ can indeed be seen as an information state transformer, with a corresponding dynamic modal operator $[\varphi]$. We now formally introduce the language with all the operators we have seen so far.

Definition 2 (Logical language of public announcements)

Given are a set of agents N and a set of atoms P . Let $p \in P$, $n \in N$, and $G \subseteq N$ be arbitrary. The language of public announcements is inductively defined as

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \psi) \mid K_n\varphi \mid C_G\varphi \mid [\varphi]\psi$$

Definition 3 (Semantics)

Given is an epistemic model $M = \langle S, \sim, V \rangle$. We define:

$$\begin{aligned} M, s \models p & : \text{iff } s \in V_p \\ M, s \models \neg\varphi & : \text{iff } M, s \not\models \varphi \\ M, s \models \varphi \wedge \psi & : \text{iff } M, s \models \varphi \text{ and } M, s \models \psi \\ M, s \models K_n\varphi & : \text{iff for all } t \in S : s \sim_n t \text{ implies } M, t \models \varphi \\ M, s \models C_G\varphi & : \text{iff for all } t \in S : s \sim_G t \text{ implies } M, t \models \varphi \\ M, s \models [\varphi]\psi & : \text{iff } M, s \models \varphi \text{ implies } M|\varphi, s \models \psi \end{aligned}$$

where $M|\varphi := \langle S', \sim', V' \rangle$ is defined as follows:

$$\begin{aligned} S' & := \{s' \in S \mid M, s' \models \varphi\} \\ \sim'_n & := \sim_n \cap (S' \times S') \\ V'_p & := V_p \cap S' \end{aligned}$$

In other words: the model $M|\varphi$ is the model M restricted to all the states where φ holds, including access between states. The interpretation of the dual $\langle\varphi\rangle$ of $[\varphi]$ will be obvious: $M, s \models \langle\varphi\rangle\psi$ if and only if $M, s \models \varphi$ and $M|\varphi, s \models \psi$. Formula φ is valid on model M , notation $M \models \varphi$, if and only if for all states s in the domain of M : $M, s \models \varphi$. Formula φ is valid, notation $\models \varphi$, if and only if for all models M (of the class of models for the given parameters of N and P): $M \models \varphi$. A proof system for this logic originates with and is proved sound and complete in [BMS98], with precursors (namely completeness results for the logic *with* announcements but *without* common knowledge) in [Pla89] and [GG97].

After Anne’s announcement that she does not have hearts, Cath knows that Anne has clubs (see Figure 5). We can verify this with a semantic computation as follows:

In order to prove that $Hexa, \clubsuit\heartsuit\spadesuit \models [\neg Hearts_a]K_c Clubs_a$, we have to show that $Hexa, \clubsuit\heartsuit\spadesuit \models \neg Hearts_a$ implies $Hexa|\neg Hearts_a, \clubsuit\heartsuit\spadesuit \models K_c Clubs_a$. As it is indeed the case that $Hexa, \clubsuit\heartsuit\spadesuit \models \neg Hearts_a$, it only remains to show that $Hexa|\neg Hearts_a, \clubsuit\heartsuit\spadesuit \models K_c Clubs_a$. The set of states that is equivalent to $\clubsuit\heartsuit\spadesuit$ for Cath, is the singleton set $\{\clubsuit\heartsuit\spadesuit\}$. So it is sufficient to show that $Hexa|\neg Hearts_a, \clubsuit\heartsuit\spadesuit \models Clubs_a$, which follows trivially from $\clubsuit\heartsuit\spadesuit \in V_{Clubs_a} = \{\clubsuit\heartsuit\spadesuit, \clubsuit\spadesuit\heartsuit\}$.

The semantics of public announcement is actually slightly imprecise. Consider what happens if in “ $M, s \models [\varphi]\psi$ if and only if $M, s \models \varphi$ implies $M|\varphi, s \models \psi$ ” the formula φ is false in M, s . In that case, $M|\varphi, s \models \psi$ is undefined, because s is now not part of the domain of the model $M|\varphi$. Apparently, we ‘informally’ use that an implication ‘antecedent implies consequent’ in the meta-language is not just true when the antecedent is false or the consequent is true, in the standard binary sense, where both antecedent and consequent are defined. But we also use that the implication is true when the antecedent is false even when the consequent is undefined. A more precise definition of the semantics of public announcement, that does not have that informality, is: $M, s \models [\varphi]\psi$ if and only if for all (M', s') such that $(M, s) \llbracket [\varphi] \rrbracket (M', s')$: $(M', s') \models \psi$. In this definition, $(M, s) \llbracket [\varphi] \rrbracket (M', s')$ holds if and only if $M' = M|\varphi$ and $s = s'$. The general definition of the interpretation of epistemic actions, of which ‘announcement’ is just an example, has a very similar form.

To give the reader a feel for what goes and what not in this logic, we give some valid principles for this logic. In all cases we only give motivation and we refrain from proofs.

If an announcement can be executed, there is only one way to do it:

$$\langle\varphi\rangle\psi \rightarrow [\varphi]\psi \text{ is valid}$$

This is a simple consequence of the functionality of the state transition semantics for the announcement. Of course, the converse $[\varphi]\psi \rightarrow \langle\varphi\rangle\psi$ does not hold. Take $\varphi = \psi = \perp$ (\perp is ‘falsum’). We now have that $[\perp]\perp$ is valid (for trivial reasons) but $\langle\perp\rangle\perp$ is, of course, always false, for the same trivial reason that no epistemic

state satisfies \perp ! Related to the functionality and partiality of ‘announcement’ are that all of the following are equivalent:

- $\varphi \rightarrow [\varphi]\psi$
- $\varphi \rightarrow \langle \varphi \rangle \psi$
- $[\varphi]\psi$

A sequence of two announcements can always be replaced by a single, more complex announcement. Instead of first saying ‘ φ ’ and then saying ‘ ψ ’ you may as well have said for the first time ‘ φ and after that ψ ’. It is expressed in

$$[\varphi \wedge [\varphi]\psi]\chi \text{ is equivalent to } [\varphi][\psi]\chi$$

This turns out to be a quite useful feature for analyzing announcements that are made with specific intentions; or, more general, conversational implicatures à la Grice. Those intentions tend to be postconditions ψ that supposedly hold after the announcement. So the (truthful) announcement of φ with the intention of achieving ψ corresponds to the announcement $\varphi \wedge [\varphi]\psi$.

For an example sequence of two announcements, consider the following announcement, supposedly made by some outsider that has full knowledge of the epistemic state (*Hexa*, $\clubsuit\heartsuit\spadesuit$) (alternatively, such an agent can be modelled as a player with the identity relation for access):

An outsider says: “The deal of cards is neither $\clubsuit\clubsuit\heartsuit$ nor $\heartsuit\clubsuit\clubsuit$.”

This is formalized as $\neg(\text{Spades}_a \wedge \text{Clubs}_b \wedge \text{Hearts}_c) \wedge \neg(\text{Hearts}_a \wedge \text{Spades}_b \wedge \text{Clubs}_c)$. Abbreviate this announcement as **one**. See Figure 6 for the result of the announcement of **one**. Observe that none of the three players Anne, Bill, and Cath know the card deal as a result of this announcement! Now imagine that the players know (publicly) that the outsider made the announcement **one** in the happy knowledge of not revealing the deal of cards to anyone! For example, he (or she) might have been boasting about his logical prowess and the players might inadvertently have become aware of that. In other words, it becomes known that the announcement **one** was made *with the intention of keeping the players ignorant of the card deal*. Ignorance of the card deal (whatever the deal may have been) can be described as some long formula that is a conjunction of eighteen parts and that starts as $\neg K_a(\text{Clubs}_a \wedge \text{Hearts}_b \wedge \text{Spades}_c) \wedge \neg K_b(\text{Clubs}_a \wedge \text{Hearts}_b \wedge \text{Spades}_c) \wedge \neg K_c(\text{Clubs}_a \wedge \text{Hearts}_b \wedge \text{Spades}_c) \wedge \dots$ and that we abbreviate as **two**. The formula **two** is false in all states (in the model resulting from the announcement of **one**) that are a singleton equivalence class for at least one player, and true anywhere else. So it’s only true in state $\clubsuit\heartsuit\spadesuit$. For the result of the announcement of **two**, see again Figure 6. Observe that in the epistemic state resulting from **two**, all players now know the card deal! So in that epistemic state **two** is false. Now what does it mean that the players have become aware of the intention of the outsider? This means that although the outsider was actually saying **one**, he was really meaning ‘one, and after that **two**’, or in other

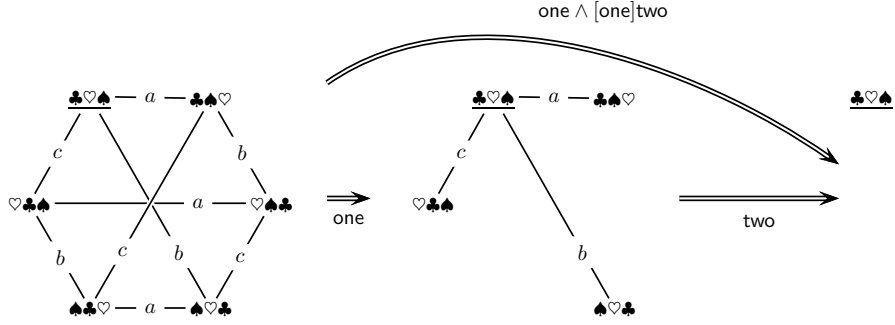


Figure 6: A sequence of two announcements can be replaced by a single announcement.

words he was saying $[\text{one}]two$. See again Figure 6. Unfortunately, we have that $Hexa, \clubsuit\heartsuit\spadesuit \models [[\text{one}]two]\neg two$. So the outsider could have kept the card deal a secret, but by intending to keep it a secret (s)he was, after all, actually revealing that secret...

The relation of the announced formula to the pre- and postconditions of the announcement is not trivial. To start with, $[\varphi]K_n\psi$ is not equivalent to $K_n[\varphi]\psi$. This is a consequence of the fact that $[\varphi]$ is a partial function. A simple counterexample is the following: in $(Hexa, \clubsuit\heartsuit\spadesuit)$ it is true that after ‘every’ announcement of ‘Anne holds hearts’, Cath knows that Anne holds clubs. This is because that announcement cannot take place in that epistemic state. In other words, we have

$$Hexa, \clubsuit\heartsuit\spadesuit \models [Hearts_a]K_c Clubs_a$$

On the other hand, it is false that Cath knows that after the announcement of Anne that she holds the hearts card (which she can imagine to take place), Cath knows that Anne holds the clubs card. On the contrary: Cath then knows that Anne holds the hearts card! So we have

$$Hexa, \clubsuit\heartsuit\spadesuit \not\models K_c[Hearts_a]Clubs_a$$

If we make $[\varphi]K_n\psi$ conditional to the executability of the announcement, an equivalence indeed holds:

$$[\varphi]K_n\psi \text{ is equivalent to } \varphi \rightarrow K_n[\varphi]\psi$$

The relation between announcement and knowledge can be formulated in various ways. One or the other may appeal more to the intuitions of the reader. Often, the ‘diamond’-versions of axioms correspond better to one’s intuitions than the ‘box’-versions. So it merely sharpens the modeller’s wits to realize that all of the following validities express the same equivalence:

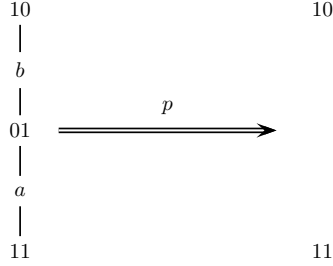


Figure 7: A state transition illustrating what a and b commonly know before and after announcement of p .

- $[\varphi]K_n\psi \leftrightarrow (\varphi \rightarrow K_n[\varphi]\psi)$
- $\langle\varphi\rangle K_n\psi \leftrightarrow (\varphi \wedge K_n(\varphi \rightarrow \langle\varphi\rangle\psi))$
- $\langle\varphi\rangle \hat{K}_n\psi \leftrightarrow (\varphi \wedge \hat{K}_n\langle\varphi\rangle\psi)$

If we restrict ourselves to the logic of announcements *without* common knowledge, every formula is logically equivalent to one in the logic without announcements. But for the logic of announcements *with* common knowledge, this is no longer the case [BMS98]. Apart from conceptual reasons, such as having a natural specification language for dynamics, *that*, one might say, is the real validation of this logical tool. Let us take a closer look at a principle relating announcements and common knowledge.

The straightforward generalization of the principle $[\varphi]K_n\psi \leftrightarrow (\varphi \rightarrow K_n[\varphi]\psi)$ relating announcement and individual knowledge would be $[\varphi]C_N\psi \leftrightarrow (\varphi \rightarrow C_N[\varphi]\psi)$. This happens to be invalid. The following countermodel M demonstrates this clearly.

Consider a model M for two agents a and b and two facts p and q . Its domain is $\{11, 01, 10\}$, where 11 is the state where p and q are both true, 01 the state where p is false and q is true, and 10 the state where p is true and q is false. Agent a cannot tell 11 and 01 apart, whereas b cannot tell 01 and 10 apart. So the partition for a on the domain is $\{11, 01\}, \{10\}$ and the partition for b on the domain is $\{11\}, \{01, 10\}$. See Figure 7.

Now consider the instance $[p]C_{ab}q \leftrightarrow (p \rightarrow C_{ab}[p]q)$ of this supposed principle. The left side of the equivalence is true in state 11 of M , whereas the right side is false in that state. We show that as follows. First, $M, 11 \models [p]C_{ab}q$ is true in 11, because $M, 11 \models p$ and $M|p, 11 \models C_{ab}q$. For the result of the announcement of p in $(M, 11)$, see Figure 7. The model $M|p$ consists of two disconnected states; obviously, $M|p, 11 \models C_{ab}q$, because $M|p, 11 \models q$ and 11 is now the only reachable state from 11.

On the other hand, we have that $M, 11 \not\models p \rightarrow C_{ab}[p]q$, because $M, 11 \models p$ but $M, 11 \not\models C_{ab}[p]q$. The last is because $11 \sim_{ab} 10$ (because $11 \sim_a 01$ and $01 \sim_b 10$), and $M, 10 \not\models [p]q$. When evaluating q in $M|p$, we are now in the *other* disconnected part of $M|p$, where q is false: $M|q, 10 \not\models q$.

Fortunately there are also other ways to get common knowledge after an announcement. The general principle is: If $\chi \rightarrow [\varphi]\psi$ and $\chi \wedge \varphi \rightarrow E_N\chi$ are valid, then $\chi \rightarrow [\varphi]C_N\psi$ is valid as well.

6 Unsuccessful updates

After announcing φ , φ may remain true but may also have become false! This will not come as a surprise to those familiar with the so-called Moore-sentences, that are already discussed in detail in the original presentation of epistemic logic in [Hin62]. This states that you cannot know that some fact is true and that you do not know that. In other words, $K(p \wedge \neg Kp)$ is inconsistent in epistemic logic. This can easily be seen by the following argument: from $K(p \wedge \neg Kp)$ follows $Kp \wedge K\neg Kp$, so follows Kp . But also, from $Kp \wedge K\neg Kp$ follows $K\neg Kp$, and from that, with ‘truthfulness’, follows $\neg Kp$. Together, Kp and $\neg Kp$ are inconsistent.

Within the setting of the logic of public announcements this can be re-described as follows: after the truthful announcement (in some given epistemic state) of $(p \wedge \neg Kp)$, this formula can no longer be true (in the resulting epistemic state). In [Ger99] this sort of announcement was called an *unsuccessful update*: you say something “because it’s true,” but unfortunately, that was not a very successful thing to do, because now it’s false!

For a different example, consider the result of Anne announcing in the epistemic state $(Hexa, \clubsuit\heartsuit\spadesuit)$: “I hold the clubs card and (at the time I am saying this) Bill does not know that”. This is an announcement of $Clubs_a \wedge \neg K_b Clubs_a$ (or of, equivalently, $K_a(Clubs_a \wedge \neg K_b Clubs_a)$; note that mixing epistemic operators for *different* agents does not make it ‘Moore’). After this announcement, Bill now knows that Anne holds the clubs card, so $K_b Clubs_a$ has become true, and therefore $\neg(Clubs_a \wedge \neg K_b Clubs_a)$ as well. The reader can simply check in Figure 8, that after this announcement the formula $\neg(Clubs_a \wedge \neg K_b Clubs_a)$ indeed holds, and therefore $Hexa, \clubsuit\heartsuit\spadesuit \models [Clubs_a \wedge \neg K_b Clubs_a]\neg(Clubs_a \wedge \neg K_b Clubs_a)$.

We appear to be deceived by some intuitive, but incorrect, communicative expectations. If an agent truthfully announces φ to a group of agents, it appears *on first sight* to be the case that (s)he ‘makes φ common knowledge’ that way: in other words, if φ holds, then after announcing that, $C_N\varphi$ holds. In other words, $\varphi \rightarrow [\varphi]C_N\varphi$ *appears* to be valid. This expectation is unwarranted, because the truth of *epistemic* (non-propositional) parts of the formula may be influenced by its announcement. On the other hand – it’s not that our intuition is *that* stupid – sometimes the expectation *is* warranted after all: the formulas that always become common knowledge after being announced, can be called

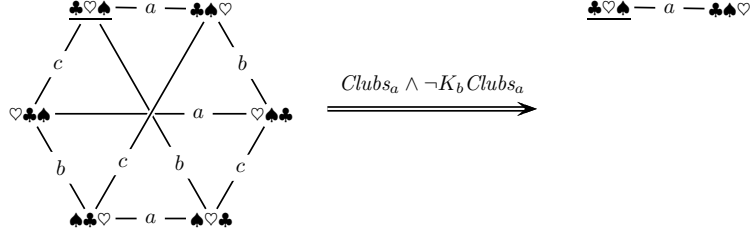


Figure 8: Anne says to Bill: “(I hold clubs and) You don’t know that I hold clubs.”

successful. What are the possibilities?

After announcing φ , φ sometimes remains true and sometimes becomes false, and this depends both on the formula *and* on the epistemic state. Consider an epistemic state for one atom p and two agents, Anne and Bill again, where Anne knows the truth about p but Bill doesn’t. This epistemic state is formally defined as $(Letter, 1)$, where the model *Letter* has domain $\{0, 1\}$, where p is true in state 1: $V_p = \{1\}$, and such that Anne can distinguish 1 from 0 but Bill cannot, so access \sim_a for a is the identity $\{(0, 0), (1, 1)\}$ and access \sim_b for b is the universal relation $\{(0, 0), (1, 1), (0, 1), (1, 0)\}$. The model is called *Letter* because it can be seen as the result of Bill seeing Anne read a *letter* which contains the truth about p . If in this epistemic state Anne says, truthfully: “I know that p ,” then after this announcement of $K_a p$ it *remains true* that $K_a p$:

$$Letter, 1 \models [K_a p] K_a p$$

This is, because in *Letter* the formula $K_a p$ is true in state 1 only, so that the model $Letter|K_a p$ consists of the singleton state 1, with reflexive access for a and b . It also becomes common knowledge that Anne knows p : we have that $Letter, 1 \models [K_a p] C_N K_a p$; although in this particular case of a singleton model, that is not very informative. We therefore also have $Letter \models K_a p \rightarrow [K_a p] C_N K_a p$ and $K_a p \rightarrow [K_a p] C_N K_a p$ is indeed valid.

But it is not always the case that announced formulas remain true. In the given epistemic state $(Letter, 1)$, Anne could on the other hand have said as well, to Bill: “You don’t know that p .” The actual implicature in this case is “Fact p is true and you don’t know that.” After this announcement of $K_a(p \wedge \neg K_b p)$, that also only succeeds in state 1, Bill knows that p , therefore $K_a(p \wedge \neg K_b p)$ is now *no longer* true

$$Letter, 1 \models [K_a(p \wedge \neg K_b p)] \neg K_a(p \wedge \neg K_b p)$$

and so it’s certainly not commonly known: $Letter, 1 \not\models [K_a(p \wedge \neg K_b p)] \neg C_N K_a(p \wedge \neg K_b p)$. So $K_a(p \wedge \neg K_b p) \rightarrow [K_a(p \wedge \neg K_b p)] C_N K_a(p \wedge \neg K_b p)$ is definitely not valid.

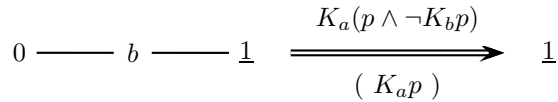


Figure 9: A simple unsuccessful update: (p and) you don't know that p . The announcement 'I know that p ' – between brackets – induces the same state transition.

The epistemic state transition induced by this announcement is visualized in Figure 9. The announcement of $K_a p$ induces the same state transition. Incidentally, like here, for a given state transition there is always a formula that induces it and *remains* true, an interesting result by Van Benthem [vB02].

In this case, we not only have that $K_a p$ remains true after being announced and that $K_a(p \wedge \neg K_b p)$ becomes false, but also that $[K_a p]K_a p$ is valid, and $[K_a(p \wedge \neg K_b p)]\neg K_a(p \wedge \neg K_b p)$ is valid. In between these extremes of 'always successful' and 'always unsuccessful' there are also formulas that sometimes remain true, and at other times – given other epistemic states – become false after an announcement. A typical example is 'not stepping forward' in the well-known Muddy Children problem [FHMV95]. The 'announcement' (implicitly, by not stepping forward) that none of the children know whether they are muddy, remains true in all epistemic states for this problem *except* the last one, in which it is an unsuccessful update: after that the muddy children know that they are muddy, and step forward. The following terminology describes all those nuances.

Definition 4 (Successful formula / Successful update)

A formula φ in the language of public announcements is *successful* if and only if $\langle \varphi \rangle \varphi$ is valid. A formula is *unsuccessful* if and only if it is not successful. Given an epistemic state (M, s) , φ is a *successful update* in (M, s) , if and only if $M, s \models \langle \varphi \rangle \varphi$; and φ is an *unsuccessful update* in (M, s) , if and only if $M, s \models \langle \varphi \rangle \neg \varphi$.

In the definitions, the switch between the 'box' and the 'diamond' versions of the announcement operator may puzzle the reader. In the definition of a *successful formula* we need the 'box'-form: $\langle \varphi \rangle \varphi$ is invalid for all φ except \top (\top stands for 'verum', 'truth'). But in the definition of a *successful update* we need the 'diamond'-form: otherwise, whenever the announcement formula is false in an epistemic state, $[\varphi]\neg\varphi$ would therefore be true, and we would be forced to call that φ an unsuccessful update. That would not capture the intuitive meaning of 'unsuccessful update', which is a property of an epistemic state transition. We must therefore assume that the announcement formula can indeed be truthfully announced. This explains the difference between the two definitions.

Announcements of (therefore true) successful formulas (the validity of $[\perp]\perp$ is considered a-typical) are always successful updates, but sometimes successful updates are on formulas that are unsuccessful. The first will be obvious: if a *successful formula* φ is true in an epistemic state (M, s) , then $\langle \varphi \rangle \varphi$ is also true

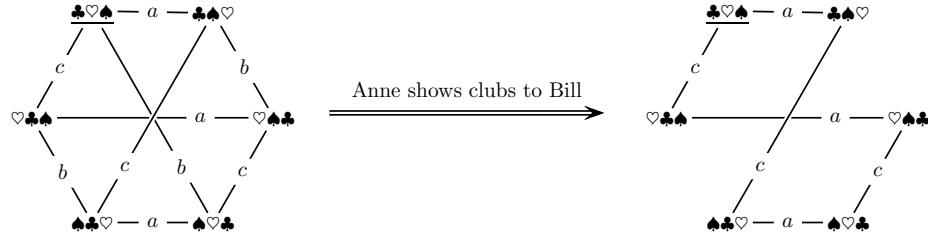


Figure 10: On the left, the Kripke model for three players each holding one card. On the right, the effect of Anne showing her clubs card to Bill.

in that state, so it is also a successful *update*. The last is, because formulas may be successful updates in one epistemic state, but unsuccessful updates in another, and from the latter follows that they are unsuccessful *formulas*.

We can link our intuitions about ‘success’ to the definition of a successful formula in a surprisingly elegant way: A formula $[\varphi]\varphi$ is valid, if and only if $[\varphi]C_N\varphi$ is valid, if and only if $\varphi \rightarrow [\varphi]C_N\varphi$ is valid. So the successful formulas ‘do what we want them to do’: if true, they become common knowledge when announced. What formulas are successful? An answer to this question is not obvious, because some inductive ways to construct the class of successful formulas fail: even if φ and ψ are successful, $\neg\varphi$, $\varphi \wedge \psi$, or $\varphi \rightarrow \psi$ may be unsuccessful. For example, both p and $\neg Kp$ are successful formulas, but, as we have seen, $p \wedge \neg Kp$ is not. A partial answer to that question and further information on unsuccessful updates, including examples, can be found in [vDK04].

7 Epistemic actions

Some epistemic actions are more complex than public announcements, where the effect of the action is always a restriction on the epistemic model. Let us reconsider the epistemic state $(Hexa, \clubsuit\heartsuit\spadesuit)$ for three players Anne, Bill and Cath, each holding one of clubs, hearts, and spades; and wherein Anne holds clubs, Bill holds hearts, and Cath holds spades. And consider again one of the example actions in the introduction:

Anne shows (only) to Bill her clubs card. Cath cannot see the face of the shown card, but notices that a card is being shown.

As always in this epistemic (and not doxastic) setting, it is assumed that it is publicly known what the players can and cannot see or hear. Call this action **showclubs**. The epistemic state transition induced by this action is visualized in Figure 10. Unlike after public announcements, in the **showclubs** action we cannot eliminate any state. Instead, all *b*-links between states have now been

severed: whatever the actual deal of cards, Bill will know that card deal and cannot imagine any alternatives. Let us show the intuitive acceptability of the resulting epistemic state. After the action *showclubs*, Anne can imagine that Cath can imagine that Anne has clubs. That much is obvious, as Anne has clubs anyway. But Anne can also imagine that Cath can imagine that Anne has hearts, because Anne can imagine Cath to have spades, and so not to know whether Anne has shown clubs or hearts; so it might have been hearts. It is even the case that Anne can imagine that Cath can imagine that Anne has spades, because Anne can imagine Cath not to have spades but hearts instead, in that case Cath would not have known whether Anne has shown clubs or spades; so it might have been spades. Note that, even though for Cath there are only two ‘possible actions’ – showing clubs or showing hearts – none of the *three* possible actions can apparently be eliminated ‘from public consideration’. The descriptions of the action *showclubs* and of the other ‘possible actions’ where Anne shows hearts or spades to Bill instead, should obviously be related: in Figure 10, this merely means shifting the point from one state to another.

But it can become even more complex. Imagine the following action, rather similar to the *showclubs* action:

Anne whispers into Bill’s ear that she does not have the spades card, given a (public) request from Bill to whisper into his ear one of the cards that she does not have.

This is the action *whispernospades*. Given that Anne has clubs, she *could* have whispered “no hearts” or “no spades”. And whatever the actual card deal was, she could always have chosen between two such options. We expect an epistemic state to result that reflects that choice, and that therefore consists of $6 \times 2 = 12$ different states. It is visualized in Figure 11. The reader may ascertain that the desirable postconditions of the action *whispernospades* indeed hold. For example, given that Bill holds hearts, Bill will now have learnt from Anne what Anne’s card is, and thus the entire deal of cards. So there should be no alternatives for Bill in the actual state (the underlined state $\clubsuit\heartsuit\spadesuit$ ‘at the back’ of the figure – for convenience, different states for the same card deal have been given the same name). But Cath does not *know* that Bill knows the card deal, as Cath can imagine that Anne actually whispered “no hearts” instead. That would have been something that Bill already knew, as he holds hearts himself – so from that action he would not have learnt very much. Except that Cath could then have imagined him to know the card deal... Note that in Figure 11 there is also another state named $\clubsuit\heartsuit\spadesuit$, ‘in the middle’ so to speak, that is accessible for Cath from the state $\clubsuit\heartsuit\spadesuit$ ‘at the back’, and that satisfies that Bill doesn’t know that Anne has clubs. Etc.

From the point of view of dynamic epistemics, a public announcement is a simple form of epistemic action: it results in a restriction of the domain. The *showhearts* action results in a refinement of accessibility relations given the same domain. The *whispernospades* action results in increased complexity of the underlying

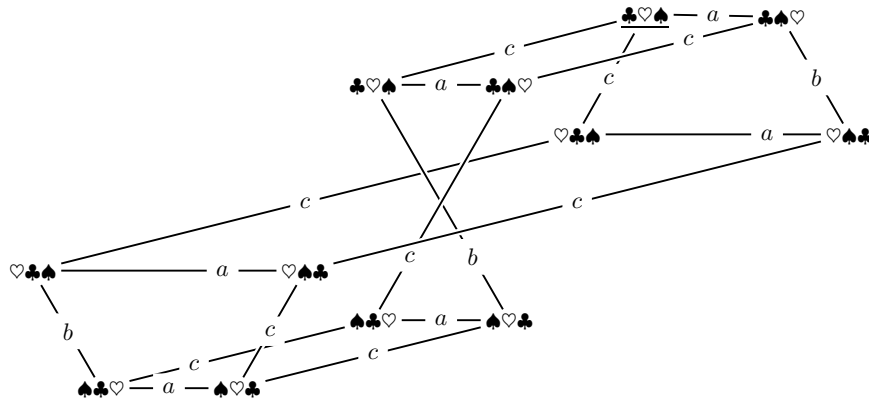


Figure 11: Anne whispers into Bill’s ear that she does not have the spades card, given a (public) request from Bill to whisper into his ear one of the cards that she does not have.

epistemic model, reflecting non-deterministic choice. To be able to model such actions a generalization of the approach used in the public announcement logic of [Pla89] was needed. Plaza’s work was more fully appreciated in the latter half of the 1990s, when subsequent, partially independent, developments took place. A stream of publications appeared around the year 2000 [GG97, Ger99, LR99b, BMS98, Bal99, vD00, vB01, BM04, vD02, tC02, Koo03, vDvdHK03]. Gerbrandy was unfamiliar with the work of Plaza at the time of his seminal publication [GG97]. It models the dynamics of introspective agents, and therefore in particular changes in belief (and, as a special case, knowledge). Its basis is a different insight into dynamics, namely along the line of work in dynamic semantics by [Vel96]. The approach in [vD00, vD02, vDvdHK03] might be called a relational action language, wherein epistemic states resulting from computing the effects of actions to subgroups (such as ‘Anne and Bill’ in the case of three cards) are used in the computations of the effects of the action for larger groups that contain that smaller group, and finally, the effects of the action for the public (such as ‘Anne, Bill, and Cath’). A different approach, and a conceptually very appealing solution, is to see a semantic action as some kind of Kripke model, an ‘action model’ so to speak, and action execution as a restricted modal product (‘the next epistemic state’) of the current epistemic state and the action model. This was first presented in [BMS98, Bal99] and its semantics recently appeared in final version in [BM04].

A crucial concept in the [vDvdHK03] approach is the ‘learn’ operator. This is a dynamic variant of the ‘common knowledge’ operator. Let’s see what it means, by paraphrasing the action `showclubs` in a way that brings this action closer to its description as an epistemic action.

Anne and Bill learn that Anne holds clubs, whereas Anne, Bill and Cath learn [that either Anne and Bill learn that Anne holds clubs, or that Anne and Bill learn that Anne holds hearts, or that Anne and Bill learn that Anne holds spades].

In other words: Anne, Bill and Cath learn that Anne and Bill learn which card Anne holds, and, actually, Anne and Bill learn that Anne holds clubs. The choice made from the three alternatives by subgroup $\{a, b\}$ is known to them only, and is hidden from c , who only knows what the three alternatives are. The description of this action in the relational action approach is

$$L_{abc}(!L_{ab}?Clubs_a \cup L_{ab}?Hearts_a \cup L_{ab}?Spades_a)$$

In this description, ‘ L ’ stands for ‘learning’, the ‘!’ indicates which of the three alternatives really happens, ‘ \cup ’ stands for non-deterministic choice, and ‘?’ stands for ‘a test on’ (the truth of the formula following it). The *whispernosspades* action is described as

$$L_{abc}(L_{ab}?¬Clubs_a \cup L_{ab}?¬Hearts_a \cup !L_{ab}?¬Spades_a)$$

Note that in this case the first option could not have been chosen, and that instead the third option has been chosen. To explain this in reasonable detail, or any of the other approaches, would carry way too far for this introduction. For details see the references.

Some rather simple actions cannot be modelled in any of the current dynamic epistemic approaches. For example, given that the action descriptions in all mentioned approaches are entirely based on the properties of the current epistemic state, one cannot distinguish between different ways in which that current state came about. Anne may *only* want to show a clubs card if some past action of Bill involved showing a spades card. But the action descriptions cannot distinguish between epistemic states that have the same (bisimilar) epistemic description but different action histories! In view of modelling game strategies, such expanded expressive power is of course essential. For another example, given the scenario where Anne receives a letter and Bill sees her reading it, suppose that the letter did not contain the truth about a single fact but contained a natural number. So instead of one fact we have infinitely many facts. Before she reads the letter, the epistemic model for that consists of infinitely many points, with universal access for both Anne and Bill, no problem at all. It is also clear what the model looks like after Anne reads the letter: Anne’s access is now the identity, and Bill’s is still the universal relation. But the action describing that Anne reads the letter, which transforms the former into the latter, has an infinitely long description, because there are infinitely many alternatives: a problem.

8 Belief, time, revision

This section presents different perspectives and other approaches. Instead of knowledge change we may want to model belief change; knowledge change can also be seen as emerging from the temporal progression of some epistemic state, using temporal and epistemic operators instead; we can see knowledge change as some kind of (deductively closed) *theory* change: a matter that has been thoroughly investigated under the header of ‘belief revision’; and there are logics that combine knowledge and belief, and degrees of belief, and probability, and changes to some or all of those.

Belief We discussed knowledge change only and not belief change – with ‘knowledge as true belief’. This was just for expository purposes. Belief change can be modelled in the same way. With the exception of the approach originating in Van Ditmarsch PhD. thesis [vD00], that so far only applies to knowledge, all mentioned approaches for dynamic epistemics only assume arbitrary accessibility relations. They therefore apply as well to structures that satisfy the properties of belief. A typical sort of epistemic action that can only be modelled in this setting is the *private announcement to a subgroup only*. Suppose that in epistemic state $(Hexa, \clubsuit\heartsuit\spadesuit)$, Anne shows Bill her clubs card, as before, but now *without* Cath noticing anything at all. In the state resulting from that action, Bill knows the card deal, as before, but Cath incorrectly believes that Bill does not know that. Such private announcements to groups are the main topic of Gerbrandy’s PhD. thesis [Ger99].

Time In temporal epistemic approaches we may express the information that Bill knows that Anne holds clubs after she said that she does not have spades, as, for example, $XK_b Clubs_a$, or $K_b^1 Clubs_a$. We then assume an underlying structure of the corresponding epistemic state transitions, for example corresponding to some such transitions in a run of an interpreted system. We cannot express the content of the action in the temporal operator. In $XK_b Clubs_a$, X is the (modal) temporal ‘next’ operator, which is interpreted as follows ‘ $XK_b Clubs_a$ is true in the current state, if in the next state (as determined by the underlying structure) $K_b Clubs_a$ is true. In $K_b^1 Clubs_a$ we do something similar, only that in this case K_b^1 is the operator describing what Bill knows at point 1 in time. Temporal epistemic logics have been fairly successful. Their computational properties are well-known and proof tools have been developed. See for example, we give just some arbitrary references here, [vdM98, DFW98, HvdMV03]. The main difference with the dynamic epistemic approach is that the temporal epistemic description takes as models systems consisting of many epistemic states together with their whole (deterministic) history and future development. Instead, in dynamic epistemics a single epistemic state – a point in that temporal structure so to speak – is sufficient: its further development is induced by the description of the action to be executed there. This may be seen as an advantage of the dynamic epistemic approach. But there are also definite advantages to the temporal epistemic approach. Consider again the Moore-sentences. After Anne

announces to Bill: “(I hold clubs and) You do not know that I hold clubs,” there is nothing inconsistent in the truth of $K_b^1(Clubs_a \wedge \neg K_b^0 Clubs)$: at point 1 in time, Bill knows that Anne holds clubs and that at point 0 in time he *did* not know that Anne holds clubs.

Belief revision In belief revision the emphasis is on theories of objective (i.e., non-epistemic) beliefs that are changed due to expansions, contractions, or revisions, typically from the point of view of a single agent. Let’s consider the point of view of Bill in ‘three cards’. In this case his ‘beliefs’ are his justified true beliefs: his knowledge. At the outset he knows that he holds hearts, but he does not know the ownership of other cards. Therefore we may assume that $Hearts_b$ is part of his set of current beliefs T . General descriptions are also part of that theory T of current beliefs, for example rules expressing that a card can only be held by a single player: exclusive disjunction of $Spades_a$, $Spades_b$, and $Spades_c$; and sentences describing single card ownership: $Hearts_b \rightarrow \neg Spades_b, \dots$; etc. Suppose the new information is ‘Anne does not hold spades’. As Bill’s current beliefs were consistent with both $Spades_a$ and $\neg Spades_a$, the belief change taking place here is an *expansion* and not a revision. The revised theory $T + \neg Spades_a$ should contain the ‘new information’ $\neg Spades_a$, and we also expect Bill to be able to derive $Clubs_a$ from that.

A general framework to describe such belief expansion in an epistemic setting, and also contractions and revisions, is given in [Seg99b]. See also [Seg99a, LR99a]. As far as the logical language is concerned, this follows more or less the following pattern:

For the example just given, Bill’s beliefs φ are described by all $K_b\varphi$ that are true in the current epistemic state. That $Hearts_b$ is part of his beliefs corresponds to the truth of $K_bHearts_b$. That both $Clubs_a$ and $\neg Clubs_a$ are absent from his beliefs, corresponds to the truth of both $\neg K_b Clubs_a$ and $\neg K_b \neg Clubs_a$ in the current state of information, before Anne’s announcement. And that $Clubs_a$ is believed by Bill after the announcement, is described by the truth of $K_b Clubs_a$ in the resulting epistemic state. The expansion with $\neg Spades_a$ corresponds to Anne’s public announcement of $\neg Spades_a$, after which $K_b \neg Spades_a$ is indeed true.

A major difference between belief revision and dynamic epistemics is that the latter, and not the former, allows higher-order belief change. In ‘three cards’ we have that from Anne’s announcement that she does not have spades, Cath does not gain any factual knowledge, but learns that Bill now knows Anne’s card. So the revision of Cath’s beliefs should involve adding a non-objective formula $K_b Clubs_a \vee K_b Hearts_a \vee K_b Spades_a$, because in the new epistemic state it is true that $K_c(K_b Clubs_a \vee K_b Hearts_a \vee K_b Spades_a)$. This general issue of updating ‘non-objective’ formulas was neglected by classical belief revision theory, partly because of complications in the form of ‘Moore’-problems. An expansion with “(I hold clubs and) You do not know that I hold clubs,” can never be successful; and ‘success’ happens to be a deeply entrenched postulate for acceptable theory revision. It was unclear how the standard AGM postulates should be generalized to include such cases.

A second important difference between dynamic epistemics and belief revision concerns not expansion but actual ‘revision’ of (possibly wrong) beliefs, i.e. updating with a formula that is inconsistent with prior beliefs. This is typically analyzed in depth by belief revision, but neglected by dynamic epistemics. Recent advances in that have been made in [vDL03, Auc03], motivated to an important extent by seminal work from Spohn [Spo88].

Suffices to give a simple example of where this comes in handy. Consider, once again, but *really* for the last time now, the action `showclubs` wherein Anne shows clubs to Bill only but with Cath noticing that. Now imagine that Cath considers it more likely that Anne shows hearts than that Anne shows clubs. And assume that Cath’s beliefs – as is common within a ‘belief revision’ setting – are determined by the things she considers most normal / most likely. With each agent we can associate a whole set of operators for all of belief, and different degrees of belief, and knowledge, and interpret these on ‘doxastic epistemic’ models, that carry a set of accessibility relations per agent. In the resulting state of information we can achieve that even though Bill knows that Anne holds clubs – $K_b Clubs_a$ – Cath believes that Bill knows that Anne holds hearts – $B_c K_b Hearts_a$. Further actions, for example Anne putting her clubs card face up on the table, then result in Cath retracting her belief in $K_b Hearts_a$ and ‘expanding’ her beliefs with $K_b Clubs_a$ instead, so we then end up with $B_c K_b Clubs_a$ again. For details, see [vD04, Auc03]. These approaches – they may incorporate infinitely many degrees of belief – also suggest overlap with achievements to combine knowledge and probability [FH94] and the dynamics of that [Koo03].

References

- [AGM85] C.E. Alchourrón, P. Gärdenfors, and D. Makinson. On the logic of theory change: partial meet contraction and revision functions. *Journal of Symbolic Logic*, 50:510–530, 1985.
- [Auc03] G. Aucher. Belief and update logic. Master’s thesis, ILLC, University of Amsterdam, Amsterdam, the Netherlands, 2003.
- [Bal99] A. Baltag. A logic of epistemic actions. In W. van der Hoek, J.-J. Meyer, and C. Witteveen, editors, *(Electronic) Proceedings of the ESSLLI 1999 workshop on Foundations and Applications of Collective Agent-Based Systems*. Utrecht University, 1999.
- [BM04] A. Baltag and L.S. Moss. Logics for epistemic programs. *Synthese*, 139:165–224, 2004. Knowledge, Rationality & Action 1–60.
- [BMS98] A. Baltag, L.S. Moss, and S. Solecki. The logic of public announcements and common knowledge. Proceedings of TARK 98, 1998.

- [DFW98] C. Dixon, M. Fisher, and M. Wooldridge. Resolution for temporal logics of knowledge. *Journal of Logic and Computation*, 8(3):345–372, 1998.
- [FH94] R. Fagin and J.Y. Halpern. Reasoning about knowledge and probability. *Journal of the Association for Computing Machinery*, 41(2):340–367, 1994.
- [FHMV95] R. Fagin, J.Y. Halpern, Y. Moses, and M.Y. Vardi. *Reasoning about Knowledge*. MIT Press, Cambridge MA, 1995.
- [Ger99] J.D. Gerbrandy. *Bisimulations on Planet Kripke*. PhD thesis, University of Amsterdam, 1999. ILLC Dissertation Series DS-1999-01.
- [GG97] J.D. Gerbrandy and W. Groeneveld. Reasoning about information change. *Journal of Logic, Language, and Information*, 6:147–169, 1997.
- [Hin62] J. Hintikka. *Knowledge and Belief*. Cornell University Press, Ithaca, NY, 1962.
- [HvdMV03] J.Y. Halpern, R. van der Meyden, and M.Y. Vardi. Complete axiomatizations for reasoning about knowledge and time. *SIAM Journal on Computing*, 2003. To appear.
- [Koo03] B.P. Kooi. *Knowledge, Chance, and Change*. PhD thesis, University of Groningen, 2003. ILLC Dissertation Series DS-2003-01.
- [LR99a] S. Lindström and W. Rabinowicz. Ddl unlimited: dynamic doxastic logic for introspective agents. *Erkenntnis*, 50:353–385, 1999.
- [LR99b] A.R. Lomuscio and M. D. Ryan. An algorithmic approach to knowledge evolution. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing (AIEDAM)*, 13(2), 1999. Special issue on Temporal Logic in Engineering.
- [MvdH95] J.-J.Ch. Meyer and W. van der Hoek. *Epistemic Logic for AI and Computer Science*. Cambridge Tracts in Theoretical Computer Science 41. Cambridge University Press, Cambridge, 1995.
- [Pla89] J.A. Plaza. Logics of public communications. In M.L. Emrich, M.S. Pfeifer, M. Hadzikadic, and Z.W. Ras, editors, *Proceedings of the 4th International Symposium on Methodologies for Intelligent Systems*, pages 201–216, 1989.
- [Seg99a] K. Segerberg. Default logic as dynamic doxastic logic. *Erkenntnis*, 50:333–352, 1999.

- [Seg99b] K. Segerberg. Two traditions in the logic of belief: bringing them together. In H.J. Ohlbach and U. Reyle, editors, *Logic, Language, and Reasoning*, pages 135–147, Dordrecht, 1999. Kluwer Academic Publishers.
- [Spo88] W. Spohn. Ordinal conditional functions: a dynamic theory of epistemic states. In W.L. Harper and B. Skyrms, editors, *Causation in Decision, Belief Change, and Statistics*, volume II, pages 105–134, 1988.
- [tC02] B. ten Cate. Internalizing epistemic actions. In Maricarmen Martinez, editor, *Proceedings of the NASSLLI-2002 student session*, Stanford University, 2002.
- [vB01] J.F.A.K. van Benthem. Logics for information update. In J.F.A.K. van Benthem, editor, *Proceedings of TARK VIII*, pages 51–88, Los Altos, 2001. Morgan Kaufmann.
- [vB02] J.F.A.K. van Benthem. One is a lonely number: on the logic of communication. Technical report, ILLC, University of Amsterdam, 2002. Report PP-2002-27 (material presented at the Logic Colloquium 2002).
- [vD00] H.P. van Ditmarsch. *Knowledge games*. PhD thesis, University of Groningen, 2000. ILLC Dissertation Series DS-2000-06.
- [vD02] H.P. van Ditmarsch. Descriptions of game actions. *Journal of Logic, Language and Information*, 11:349–365, 2002.
- [vD04] H.P. van Ditmarsch. Prolegomena to dynamic belief revision. Manuscript, 2004.
- [vDK04] H.P. van Ditmarsch and B.P. Kooi. The secret of my success. To appear in *Synthese*, 2004.
- [vDL03] H.P. van Ditmarsch and W.A. Labuschagne. A multimodal language for revising defeasible beliefs. In E. Álvarez, R. Bosch, and L. Villamil, editors, *Proceedings of the 12th International Congress of Logic, Methodology, and Philosophy of Science (LMPS)*, pages 140–141. Oviedo University Press, 2003.
- [vdM98] R. van der Meyden. Common knowledge and update in finite environments. *Information and Computation*, 140(2), 1998.
- [vDvdHK03] H.P. van Ditmarsch, W. van der Hoek, and B.P. Kooi. Concurrent dynamic epistemic logic. In V.F. Hendricks, K.F. Jorgensen, and S.A. Pedersen, editors, *Knowledge Contributors*, pages 45–82, Dordrecht, 2003. Kluwer Academic Publishers. *Synthese Library* Volume 322.

- [vDvdHK04] H.P. van Ditmarsch, W. van der Hoek, and B.P. Kooi. Dynamic epistemic logic. Manuscript, 2004.
- [Vel96] F. Veltman. Defaults in update semantics. *Journal of Philosophical Logic*, 25:221–261, 1996.