Dynamic Epistemic Logic and Knowledge Puzzles

H.P. van Ditmarsch[1,*], W. van der Hoek[2], and B.P. Kooi[3]

[1] University of Otago, New Zealand
`hans@cs.otago.ac.nz`
[2] University of Liverpool, United Kingdom
`wiebe@csc.liv.ac.uk`
[3] University of Groningen, Netherlands
`b.p.kooi@rug.nl`

Abstract. We briefly give an overview of Dynamic Epistemic Logic (DEL), mainly in semantic terms. We focus on the simplest of epistemic actions in DEL, called public announcements. We also sketch the effect of more complex epistemic actions, and briefly show how als factual change can be modelled in the same framework. We then apply the logic of public announcements in DEL to the analysis of a knowledge puzzle, called 'What Sum'.

Keywords: multiagent systems, epistemic logic, dynamic epistemic logic, belief revision.

1 Dynamic Epistemic Logic

- A: As you know, ICCS 2007 takes place in Sheffield.
- B: Is that so? When I google ICCS 2007, I end up in Beijing: `http://www.iccs-meeting.org/`, where the International Conference on Computational Science 2007 is organized.
- C: Now is ICCS in Sheffield, or not?
- A: Ah, but I am talking about the 15th international conference on conceptual structures, `http://www.iccs.info/`, something entirely different from the Beijing thing! And it certainly takes place in Sheffield.
- C: OK. We now all know that the conceptual structures conference is indeed in Sheffield.
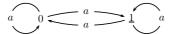
1.1 Single-Agent Knowledge

One way of resolving an uncertainty is being explicit about acronyms. In general, given uncertainty about the truth of a propositional statement 'ICCS is in Sheffield', abbreviated as Sheff, and from now referring to the conceptual structures conference, one can associate two different 'worlds', or 'states' with it: one
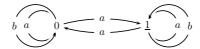
---

wherein it is true, and one wherein it is false. The reader of course knows it is true. But an agent a who cannot distinguish the actual state where it is true from the state where it is false, is said not to *know* that Sheff. Knowing a statement to be true means that in all states that one considers possible, that statement is true. As, according to a's information, there is a possible state where Sheff is true and one where it is false, Sheff is therefore not known. We write $\neg K_a$Sheff for that; K_a stands for 'agent a knows'. A simple and abstract representation of that situation, called *epistemic model* or *information state*, (formal definitions of language and semantics will be given in Section 2.1) is

The two states are named 0 and 1. In state 0 the proposition Sheff is false, and in state 1 it is true. The arrows, labelled with a, stand for what agent a considers possible. First, we reason from the actual state of affairs: ICCS is in Sheffield. This is true in state 1. To indicate that 1 is the actual state, we have underlined it in the figure. In state 1, agent a, *Anne*, considers it possible that Sheff is true. Therefore there is a reflexive arrow from 1 to itself. But she also considers it possible that Sheff is false. Therefore there is an arrow from 1 to 0. There are more arrows in the picture! The reason for this extra structure is that we assume that Anne is aware of her ignorance of Sheff. In terms of our simplifying assumptions such awareness can be equated with knowledge. We then get: Anne knows that she does not know that ICCS is in Sheffield. Formally, this is: $K_a \neg K_a$Sheff. Such reflection is called a higher-order aspect of knowledge. In practice this means stacking K-operators to greater depth than just one. To represent this higher-order knowledge, Anne should be able to reason from the perspective that the actual state *were* 0, where Sheff is false. From that perspective, it is also both conceivable that Sheff is true and that Sheff is false. Therefore, there are also a-arrows from 0 to 1 and from 0 to itself.

1.2   Multi-Agent Knowledge

Let us introduce a second agent b, say *Bill*. We can model that Bill knows that Anne is ignorant of Sheff but he himself is aware of the truth, and that even Anne knows that. For a more realistic setting than the initial one, imagine the truth about Sheff to be written on a sheet of paper in a closed envelope that is handed to Bill, in the presense of Anne, by some third party stating 'this envelope contains the truth about Sheff,' after which Bill opens the envelope and reads its contents still in the presence of Anne. (Epistemic logic is full of such scenarios.) The resulting situation is now perfectly modelled by the above assumptions. And in fact by something stronger: Bill knows that Anne knows that Bill knows the truth about Sheff, and Anne knows that, etc. We say that Anne and Bill have *common knowledge* about the situation where Anne is ignorant and Bill knowledgeable about Sheff: $C_{ab}[\neg(K_a$Sheff $\lor K_a \neg$Sheff$) \land (($Sheff $\land K_b$Sheff$) \lor (\neg$Sheff $\land K_a \neg$Sheff$))]$, depicted as:

Common knowledge formalizes what agents in a group know about each other, and about each other's knowledge, and so on ad infinitum. We can also think of common knowledge as *background knowledge* describing the structure of information states. In that sense, the model above is precisely described by the common knowledge formula we gave. This aspect will reappear when we present knowledge puzzles.

The *epistemic logic* we have now used to formalize knowledge, and how to interpret this in relational structures as above, is generally said to have started with [8]. The aspect of knowledge iterations, and how the concept of common knowledge formalizes arbitrary finite iterations, is from somewhat later date. The standard reference is [10] and another early source is within economics, [3]. Recent introductions into epistemic logic are [6,13].

1.3   Public Announcements

What happens if the truth about Sheff becomes known to Anne, for example by way of a public announcement of Sheff, or by Bill saying "I know that ICCS is in Sheffield"? We consider the first, although in this example the informational consequences of both are the same. After the announcement of Sheff, both Anne and Bill now know that Sheff is true, and that the other knows, ad infinitum. Structurally, the result of a public announcement is the restriction of the information state to those states where the announcement is true, and such that all arrows are kept between these remaining states. We get

On the right it is common knowledge that Sheff: C_{ab}Sheff. From this also follows that Anne knows Sheff: K_aSheff. In the information state before the announcement, on the left, Sheff is true, and $\neg K_a$Sheff is true. We can also express the dynamic effect of the announcement in the logical language, by stating that initially (on the left) $\neg(K_a$Sheff $\vee K_a\neg$Sheff$) \wedge [$Sheff$]K_a$Sheff is true, for 'Anne does not know whether ICCS is in Sheffield and ('but') after the announcement that ICCS is in Sheffield, she knows that ICCS is in Sheffield." The operator K_a is called an epistemic operator, and the operator [Sheff] is called a dynamic operator. It is interpreted by means of a transition from one information state to another one, as above: to interpret [Sheff]K_aSheff on the left, first do the transition to the right, and interpret K_aSheff there. As this is true, so is, on the left, [Sheff]K_aSheff.

Standardly accepted properties of knowledge are that: known information is true ($K\varphi \to \varphi$, for all φ), you are aware of ('know') your knowledge ($K\varphi \to$

$KK\varphi$), and you are aware of your ignorance ($\neg K\varphi \to K\neg K\varphi$). These correspond to the structural properties of reflexivity, transitivity, and euclidicity. We do not explain this in detail, because the outcome is that together these properties ensure that the indistinguishability relation above is always an *equivalence relation*. We can therefore think of the domain of states as being partitioned into *epistemic classes*, such that every class consists of epistemically indistinguishable states for each agent. For equivalence relations, a simpler visualization is sufficient, where we link states in the same class, and where singleton classes, consisting of one state, are implicit. We then get:

$$0 \; \overline{\quad\quad} \; a \; \overline{\quad\quad} \; \underline{1} \qquad\qquad \xrightarrow{\text{Sheff}} \qquad\qquad \underline{1}$$

We will extensively use this visualization in the next section. The dynamics so far comes under the name of 'public announcement logic'. Standard references are [16,4,7,23].

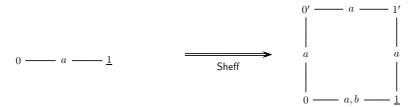## 1.4   Sentences That Become False Because They Are Announced

In public announcement logic, not all formulas remain true after their announcement, in other words, $[\varphi]\varphi$ is *not* a principle of the logic. Some formulas involving epistemic operators become **false** after being announced! Given the information state again wherein Anne is ignorant about Sheff but Bill not, consider Bill saying to Anne: "ICCS is in Sheffield but you don't know that." This is formalized as $K_b(\text{Sheff} \wedge \neg K_a\text{Sheff})$ (the initial K_b can be equated with 'Bill says (truthfully)'). Following the same recipe of restricting the information state to those of its elements where the new information is true, we again get

$$0 \; \overline{\quad\quad} \; a \; \overline{\quad\quad} \; \underline{1} \qquad\qquad \xrightarrow{K_b(\text{Sheff} \wedge \neg K_a\text{Sheff})} \qquad \underline{1}$$

In the resulting structure, Anne knows that ICCS is in Sheffield: $K_a\text{Sheff}$. Therefore the announced formula $K_b(\text{Sheff} \wedge \neg K_a\text{Sheff})$, that was true before the announcement, has become false after the announcement. One can also say that the statement about Anne's ignorance led her to factual knowledge (namely of Sheff). In the somewhat different setting that formulas of the form $p \wedge \neg K_n p$ cannot be consistently known, this phenomenon is called the Moore-paradox [14,8]. In the underlying dynamic setting it has been described as an *unsuccessful update* [7,23,17]. Similarly, statements about ignorance in the knowledge puzzles to be discussed next, may lead to factual knowledge about the numbers these puzzles are dealing with.

1.5   More Complex Dynamics

More complex informative scenarios are also conceivable. A variation of the scenario where a closed envelope containing Sheff or ¬Sheff is handed to Bill, is that, after the delivery, Bill is out of Anne's sight for a moment such that, when she returns, Bill *may have* quickly opened the envelope and read its contents. But Anne
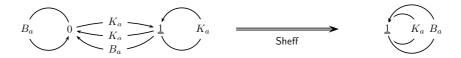
is not sure whether Bill has done that, or not (and again, we assume some common awareness of this scenario between Anne and Bill). Now, even given that Sheff is true, there are two outcomes, that are indistinguishable for Anne: either Bill opened the envelope, and now knows whether Sheff, or he didn't, in which case he remains ignorant. This informational transition can be depicted as

$$0 \xrightarrow{\quad a \quad} \underline{1} \qquad\xRightarrow{\text{Sheff}}\qquad
\begin{array}{ccc}
0' & \!\!\!\!\underline{\quad\ a\ \quad}\!\!\!\! & 1' \\
| & & | \\
a & & a \\
| & & | \\
0 & \!\!\!\!\underline{\quad a,b\quad}\!\!\!\! & \underline{1}
\end{array}$$

wherein in both 0 and $0'$, Sheff is false and in both 1 and $1'$ Sheff is true. In words: state 0 is the result when Sheff being false and b not reading the contents of the enveloppe, $0'$ is the result when Sheff is false but b *does* read this. We assume (but do not draw) transitivity of indistinguishability relations, such that 0 and $1'$, and $0'$ and $\underline{1}$, are also indistinguishable for Anne. In fact none of the four states can now be distinguished by her. Note that the above represents that Bill actually did *not* look at the contents for the envelope, as in the underlined state he still considers the alternative where Sheff is false! Such more complex informative actions have been investigated by [4,7,19] and are also a major topic in [23].

1.6   Belief and Plausible Reasoning

The difference between knowledge and belief is that beliefs may be false, whereas knowledge is supposed to be true. Imagine that we do not have two agents Anne and Bill, but just Anne. One way to model tentative belief is to associate that with preferences in a structure. For example, given the states 0 and 1 where Sheff is false and true, Anne considers it more likely that it is false: she prefers 0 over 1. But given just state 1 wherein Sheff it true, she prefers 1 (over nothing, given the absence of other states). Using labels K_a to denote epistemic indistinguishabity and B_a doxastic preference, we can depict the transition from Anne believing that Sheff is false to Anne believing that Sheff is true, as

$$B_a \overset{\curvearrowright}{\underset{\curvearrowleft}{0}} \underset{\substack{K_a \\ B_a}}{\overset{K_a}{\rightleftarrows}} \underset{}{\overset{\curvearrowright}{1}} K_a \qquad\xRightarrow{\text{Sheff}}\qquad \overset{\curvearrowleft}{\underset{\curvearrowright}{1}} \; K_a \; B_a$$

The B_a-arrow on the right, the was not there on the left, reflects that Anne now prefers 1. The belief operator is written as B. On the left, Anne does not know that Sheff, and even believes that Sheff is false: $\neg K_a\text{Sheff} \wedge B_a\neg\text{Sheff}$.

Whereas on the right, she knows (and therefore also believes) that Sheff is true: $K_a\mathsf{Sheff} \wedge B_a\mathsf{Sheff}$. Such a transition is remarkable because her factual beliefs are now the opposite from what they were before. As this is a typical operation 'revision' in the area indeed known as (AGM) belief revision [1], these settings allow for what is known as dynamic (higher-order) belief revision, on which fairly recently progress has been made [2,20,5].

1.7   Factual Change

Let us consider a different scenario involving Anne (a). Let Crack stand for 'the Ming vase over there has a crack.' An does not know whether Crack is true: she does not see one, but it may be at the back of the vase, from her perspective. Now Barteld walks past, brushes against the vase, and it falls to the ground where it shatters in a thousand pieces. The uncertainty is removed: it is now most certainly cracked. Not by someone looking at the back of the vase, leaving the state of the world unchanged, but by someone changing the world. Given ignorant Anne and knowledgeable Bill again, who observe Barteld's antics, this action can again be described as an information state transition; if we let 1 stand for 'Crack is true, as:

$$0 \longrightarrow a \longrightarrow \underline{1} \qquad \xRightarrow{\mathsf{Crack}} \qquad 1 \longrightarrow a \longrightarrow \underline{1}$$

Note that Bill is omniscient: all his equivalence classes are singletons. More significantly, observer that the state 0 on the left hand side transforms in a state 1: the truth of Crack changes! Again, it is now common knowledge that Crack is true. In this case, we can write this as $\neg K_a\mathsf{Crack} \wedge [\mathsf{Crack} := \top]K_a\mathsf{Crack}$: initially, Anne did not know whether the vase had a crack but after the vase shattered (an *assignment*), she knows: it is now true in *both* states she considers possible. Informative and factual changes can also be combined into more complex actions. Such matters are being investigated in [22,18,9].

  This ends our sweeping overview of dynamic epistemic logic. We now proceed to the analysis of a knowledge puzzle. We can do this without the formalities of the logical language and semantics, and focus on structural transitions for information states as the above. But let it be known that this formal level certainly hovers at the back of anything we playfully introduce! Indeed, such formalisations are behind various model checking operations that can be performed to verify our less formal statements below [21].

2   Knowledge Puzzles

The following riddle (transcribed in our terminology) appeared in Math Horizons in 2004, as 'Problem 182' in a regular problem section of the journal [11].

> *Each of agents Anne, Bill, and Cath has a positive integer on its forehead. They can only see the foreheads of others. One of the numbers is the sum of the other two. All the previous is common knowledge. The agents now successively make the truthful announcements:*

    i. *Anne: "I do not know my number."*
   ii. *Bill: "I do not know my number."*
  iii. *Cath: "I do not know my number."*
  iv. *Anne: "I know my number. It is 50."*
What are the other numbers?

You know your own number if and only if you know which of the three numbers is the sum. We therefore call the riddle: 'What Sum'. It combines features from wisemen or Muddy Children puzzles [15] with features from the Sum and Product riddle [12]. A common feature in such riddles is that we are given a multi-agent system, and that successive announcements of ignorance finally result in its opposite, typically factual knowledge. In a global state of such a system [6] each agent or processor has a local state, and there is common knowledge that each agent only knows its local state, and what the extent is of the domain. If the domain consists of the full cartesian product of the sets of local state values, it is common knowledge that agents are ignorant about others' local states. In that case an ignorance announcement has no informative value. For ignorance statements to be informative, the domain should be more restrictive than the full cartesian product (here: we only have triples (x, y, z) in which one number is the sum of the other two). As in Muddy Children, we do not take the 'real' state of the agent (the number on its forehead) as its local state, but instead the information seen on the foreheads of others (the other numbers). 'Sum and Product' is also about numbers, and even about sums of numbers, and the announcements are similar. (However, the structure of the background knowledge is very different: in 'Sum and Product', there are only three distinct *abc*-classes [23].)

Other epistemic riddles involve cryptography and the verification of information security protocols ('Russian Cards', see [23]), or involve communication protocols with private signals involving diffusion of information in a distributed environment ('100 prisoners and a lightbulb', see [25]). The understanding of such riddles is facilitated by the availability of suitable specification languages. For 'What Sum' we propose the logic of public announcements. But '100 prisoners and a lightbulb' requires the complex dynamics and factual change, as discussed in Subsections 1.5 and 1.7. Verification tools, such as DEMO, an epistemic model checker [24], can be used when analyzing such puzzles.

2.1 Details on Public Announcement Logic

We now formally reintroduce the language of public announcement logic, the epistemic structures in which it can be interpreted, and sufficient details of the semantics to understand epistemic statements and transitions caused by announcements. Given a finite set of agents N and a finite or countably infinite set of atoms P, the language of public announcement logic is inductively defined as

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

where $p \in P$, $n \in N$, and $B \subseteq N$ are arbitrary. Other propositional and epistemic operators are introduced by abbreviation. For $K_n\varphi$, read 'agent n knows formula

φ'. For example, if Anne knows that her number is 50, we can write $K_a 50_a$, where a stands for Anne and some set of atomic propositions is assumed that contains 50_a to represent 'Anne has the number 50.' For $C_B\varphi$, read 'group of agents B commonly know formula φ'. For example, we have that $C_{abc}(20_b \rightarrow K_a 20_b)$: it is common knowledge to Anne, Bill, and Cath, that if Bill's number is 20, Anne knows that (because she can see Bill's number on his forehead)—instead of $\{a, b, c\}$ we often write abc. For $[\varphi]\psi$, read 'after public announcement of φ, formula ψ (is true)'. "After Anne announces (I know my number. It is 50.) it is common knowledge that Bill's number is 20" is formalised as $[K_a 50_a]C_{abc}20_b$.

The basic structure is an epistemic model, which was already informally used in the previous section. It is a Kripke model, wherein all accessibility relations are equivalence relations. An *epistemic model* $M = \langle S, \sim, V \rangle$ consists of a *domain* S of (factual) *states* (or 'worlds'), *accessibility* $\sim : N \rightarrow \mathcal{P}(S \times S)$, where each $\sim_n$ is an equivalence relation, and a *valuation* $V : P \rightarrow \mathcal{P}(S)$. For $s \in S$, (M, s) is an *epistemic state*. Given two states s, s' in the domain, $s \sim_n s'$ means that s is indistinguishable from s' for agent n on the basis of its information. For example, at the beginning of the riddle, triples $(2, 14, 16)$ and $(30, 14, 16)$ are indistinguishable for Anne but not for Bill nor for Cath. Therefore, assuming a domain of natural number triples, we have that $(2, 14, 16) \sim_a (30, 14, 16)$. Given a state s, Anne knows a statement if it is true in all states she renders indistinguishable from s: given $(2, 14, 16)$, Anne knows that Bill has 14, $K_a 14_b$, because Bill has 14 in both $(2, 14, 16)$ and $(30, 14, 16)$, the two states that she cannot distinguish. The group accessibility relation $\sim_B$ is the transitive and reflexive closure of the union of all accessibility relations for the individuals in B: $\sim_B \equiv (\bigcup_{n \in B} \sim_n)^*$. This relation is used to interpret common knowledge for group B. Instead of '$\sim_B$ equivalence class' ($\sim_n$ equivalence class) we write B-class (n-class). For example, if Anne sees two even numbers she knows that it is common knowledge that all numbers are even (a full description would be cumbersomely long in this case).

The dynamic modal operator $[\varphi]$ is interpreted as an epistemic state transformer. Announcements are assumed to be truthful, and this is commonly known by all agents. It results in the restriction of the domain to all states where the announcement is true, retaining all uncertainty on that restricted domain. We will only present this informally in the subsequent analysis, by way of such resulting restrictions.

2.2 Formalisation of 'What Sum'

The set of agents $\{a, b, c\}$ represent Anne, Bill and Cath, respectively. Atomic propositions i_n represent that agent n has natural number i on its forehead. Therefore the set of atoms is $\{i_n \mid i \in \mathbb{N}^+ \text{ and } n \in \{a, b, c\}\}$. If Anne sees (knows) that Bill has 20 on his forehead and Cath 30, we describe this as $K_a(20_b \wedge 30_c)$. If an upper bound max for all numbers *were* specified in the riddle, the number of states would be finite and "knowing the others' numbers" would be described as a disjunction $\bigvee_{y,z \leq \mathsf{max}} K_a(y_b \wedge z_c)$. In 'What Sum' no upper bound is given, so strictly we now have an *infinitary* disjunction $\bigvee_{y,z \in \mathbb{N}^+} K_a(y_b \wedge z_c)$, such that

Anne saying: "I don't know my number" is similarly described as $\neg \bigvee_{x \in \mathbb{N}^+} K_a x_a$ (or $\bigwedge_{x \in \mathbb{N}^+} \neg K_a x_a$). Infinitary descriptions are, unlike infinitely large models, not permitted in this (propositional) logic. For the moment, therefore, we restrict our analysis of the updates mainly to a semantic one.

The epistemic model $\mathcal{T} = \langle S, \sim, V \rangle$ is defined as follows ($x, y, z \in \mathbb{N}^+$):

$$S \equiv \{(x, y, z) \mid x = y + z \text{ or } y = x + z \text{ or } z = x + y\}$$

$$\forall (x, y, z), (x', y', z') \in S \ (x, y, z) \sim_a (x', y', z') \ \text{ iff } \ y = y' \text{ and } z = z'$$
$$\forall (x, y, z), (x', y', z') \in S \ (x, y, z) \sim_b (x', y', z') \ \text{ iff } \ x = x' \text{ and } z = z'$$
$$\forall (x, y, z), (x', y', z') \in S \ (x, y, z) \sim_c (x', y', z') \ \text{ iff } \ x = x' \text{ and } y = y'$$
$$\forall k \in \mathbb{N}^+ \ \ V(k_a) = \{(x, y, z) \in S \mid x = k\}$$
$$\forall k \in \mathbb{N}^+ \ \ V(k_b) = \{(x, y, z) \in S \mid y = k\}$$
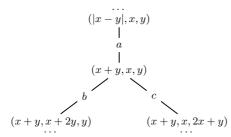$$\forall k \in \mathbb{N}^+ \ \ V(k_c) = \{(x, y, z) \in S \mid z = k\}$$

For the sequel, it is important to realise that, for each agent n, the set of states in which (s)he does *not* know his or her own numbers, is

$$\{(x, y, z) \in S \ \mid \ \exists (x', y', z') \neq (x, y, z) \text{ and } (x, y, z) \sim_n (x', y', z')\}$$

A relevant question is what the background knowledge is that is available to the agents, i.e., what the *abc*-classes in the model are (an *abc*-class, or $\{a, b, c\}$ equivalence class, of a state s in the model consists of all states t such that $s \sim_{\{a,b,c\}} t$, where $\sim_{\{a,b,c\}} = (\sim_a \cup \sim_b \cup \sim_c)^*$, as above).

An *abc*-class in $\mathcal{T}$ can be visualised as an infinite binary tree. The depth of the tree reflects the following order on number triples in the domain of $\mathcal{T}$: $(x, y, z) > (u, v, w)$ iff ($x > u$ and $y = v$ and $z = w$) or ($x = u$ and $y > v$ and $z = w$) or ($x = u$ and $y = v$ and $z > w$). If $(x, y, z) > (u, v, w)$ according to this definition, (x, y, z) is a child of (u, v, w) in the tree. Every node except the root has one predecessor and two successors, as in Figure 1.

The root of each tree has label $(2x, x, x)$ or $(x, 2x, x)$ or $(x, x, 2x)$. In each such a root, at least one of the agents knows his or her own number. The idea now is that after an announcement by agent n of the form "I don't know my number", such a root where n knows his number can be eliminated, and the tree gets split in two subtrees. Starting with an arbitrary state (x, y, z) in the tree, such that one is the *sum* of the other two, replace that sum by the *difference* of the other two; one of those other two has now become the sum; if you repeat the procedure, you *always* end up with two equal numbers and their sum. An agent who sees two equal numbers, immediately infers that its own number must be their sum (twice the number that is seen), because otherwise it would have to be their difference 0 which is not a positive natural number. It will be obvious that: the structure truly is a forest (a set of trees), because each node only has a single parent; all nodes except roots are triples of three *different* numbers; and all trees are infinite. All *abc*-trees are isomorphic modulo (*i*) a multiplication factor for the numbers occurring in the arguments of the node labels, and modulo (*ii*) a permutation of arguments and a corresponding swap of agents, i.e., swap of arc labels. For example, the numbers occurring in the tree with root $(6, 3, 3)$ are thrice the

As an illustration, consider $(30,14,16)$ again. It has the following path to the root: $(30,14,16) \sim_a (2,14,16) \sim_c (2,14,12) \sim_b (2,10,12) \sim_c (2,10,8) \sim_b (2,6,8) \sim_c (2,6,4) \sim_b (2,2,4)$. The only sibling of $(30,14,16)$ is $(2,18,16)$: it is $\sim_b$-connected to the shared parent $(2,14,16)$.

Fig. 1. (Left) Modulo agent symmetry, all parts of the model $\mathcal{T}$ branch as here. Arcs connecting nodes are labelled with the agent who cannot distinguish those nodes. (Right) A description of part of the tree with $(30,14,16)$.

corresponding numbers in the tree with root $(2,1,1)$; the tree with root $(2,1,1)$ is like the tree for root $(1,2,1)$ by applying permutation (213) to arguments and (alphabetically ordered) agent labels alike. The left side of Figure 3 shows the trees with roots $(2,1,1)$, $(1,2,1)$, and $(1,1,2)$. For simplicity, we write 211 instead of $(2,1,1)$, etc. In the left tree, for Bill $(2,1,1)$ is indistinguishable from $(2,3,1)$ wherein his number is the sum of the other two instead of their difference; for Anne triple $(2,3,1)$ is indistinguishable from $(4,3,1)$, etc.

Processing Announcements. The result of an announcement is the restriction of the model to all states where the announcement is true. We can also apply this to the ignorance announcements of agents in 'What Sum'. Consider an *abc*-tree T in $\mathcal{T}$. Let n be an arbitrary agent. Either the root of T is a singleton n-class, or all its n-classes consist of two elements: a two-element class represents the agent's uncertainty about its own number. An ignorance announcement by agent n in this riddle corresponds to removal of all singleton n-classes from the model $\mathcal{T}$. This means that *some* of the model's trees are split into two subtrees (with both children of the original root now roots of infinite trees).

An ignorance announcement may have very different effects on *abc*-classes that are the same modulo agent permutations. For example, given *abc*-classes in $\mathcal{T}$ with roots 121, 112, and 211, the effect of Anne saying that she does not know her number *only* results in elimination of 211, as only the first *abc*-class contains an *a*-singleton. Given 211, Anne knows that she has number 2 (as 0 is excluded). But triple 112 she cannot distinguish from 312, and 121 not from 321. Thus one proceeds with all three announcements. See also Figure 2.

Solving the riddle. We have now sufficient background to solve the riddle. We apply the successive ignorance announcements to the three classes with roots $(2,1,1)$, $(1,2,1)$, and $(1,1,2)$, determine the triples wherein Anne knows the numbers, and from those, wherein Anne's number divides 50. See Figure 3—note that in triple $(8,3,5)$ Anne also knows her number: the alternative $(2,3,5)$ wherein her number is 2 has been eliminated by Cath's, last, ignorance announcement. The *unique* triple wherein Anne's number divides 50 is $(5,2,3)$. In other words, the unique *abc*-tree in the *entire* model $\mathcal{T}$ where Anne knows that she
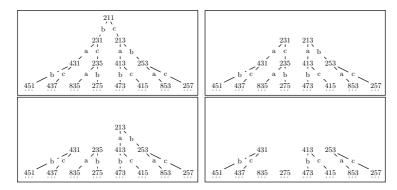
Fig. 2. The results of three ignorance announcements (by a, b, and c, respectively) on the abc-class with root $(2, 1, 1)$
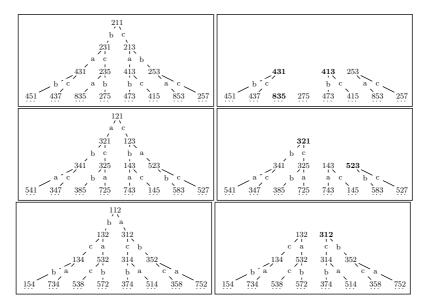

Fig. 3. On the left, abc-classes of the model $\mathcal{T}$ with root 211, 121, and 112. Any other abc-class is isomorphic to one of these, modulo a multiplication factor. The results of the (combined) three ignorance announcements on those abc-classes are on the right. The triples in bold are those where Anne knows her number.

has 50 after the three ignorance announcements, is the one with root $(10, 20, 10)$. The solution to the riddle is therefore that Bill has 20 and Cath has 30. After the three announcements in the abc-class with root $(10, 20, 10)$, the triple $(50, 20, 30)$ remains wherein Anne knows that Bill has 20 and Cath 30.

The original riddle could have been more restrictive: in the quoted version [11] it is *not* required to determine who holds which other number, but as we

have seen this can also be determined. It also occurred to us that the original riddle has the following attractive variant:

> *Each of agents Anne, Bill, and Cath has a positive integer on its forehead. They can only see the foreheads of others. One of the numbers is the sum of the other two. All the previous is common knowledge. The agents now successively make the truthful announcements:*
> *i. Anne: "I do not know my number."*
> *ii. Bill: "I do not know my number."*
> *iii. Cath: "I do not know my number."*
> *What are the numbers, if Anne now knows her number and if all numbers are prime?*

Consulting Figure 3, it will be obvious that the answer should be: '5, 2, and 3'.

2.3   Variations of the Riddle

Suppose we use an upper bound max for the numbers. Let $\mathcal{T}^{\mathsf{max}}$ be the corresponding epistemic model. An abc-tree is now cut at the depth where nodes (x, y, z) occur such that the sum of two of the arguments x, y, z exceeds max. This finite approximation may not seem a big deal but it makes the problem completely different: abc-classes will not just have *roots* wherein the agent may know his number (because the other numbers are equal) but will also have *leaves* wherein the agent may know his number (because the sum of the other two numbers exceeds max). In other words, we have far more singleton equivalence classes. Let $\mathsf{max} = 10$. Node $(2, 5, 7)$ in the abc-class with root $(2, 1, 1)$ has only a b-child $(2, 9, 7)$ and a c-parent $(2, 5, 3)$, and not an a-child, as $5 + 7 = 12 > \mathsf{max}$. So Anne immediately knows that her number is 2. All roots $(2x, x, x)$ with $3x > \mathsf{max}$ form singleton abc-classes in $\mathcal{T}^{\mathsf{max}}$, for the same reason. In such models it is no longer the case that all equivalence classes are isomorphic modulo a multiplication factor and swapping of agent labels—but sufficient good properties for systematic exploration remain (see [21] for details).

Suppose we start counting from 0 instead of 1. In that case each abc-equivalence class with root $(2x, x, x)$ is extended with one more node: the new root $(0, x, x)$ is indistinguishable from $(2x, x, x)$ for Anne. An agent who sees a 0, infers that his number must be the other number that (s)he sees. If there is a 0, two of the three agents see that. Therefore, the root has just one child $(2x, x, x)$; if the triple is $(0, x, x)$ Bill and Cath know that their number is x.

Now consider the following version of the riddle: If the three numbers have an upper bound, and if 0 is allowed, for which range of the upper bound does Anne now *always* know the numbers after the three announcements? Let max be the upper bound. The requested range includes $\mathsf{max} = 10$. Figure 4 shows that from abc-class with root 011 the triples 211 and 213 remain. There is one other abc-class in the epistemic model $\mathcal{T}_0^{10}$ (for $0 \le x, y, z \le 10$) that remains non-empty after the three announcements, namely the one with root 022: the triples 242 and 246 then remain. Therefore, whatever the numbers, Anne now knows her's.
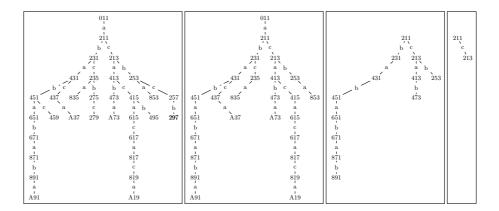
Fig. 4. Successive announcements in the *abc*-class with root 011 in model $\mathcal{T}_0^{10}$. The horizontal order of branches has no meaning. Symbol A represents 10.

The solution range is: $8 \leq \mathsf{max} \leq 13$. This means that if $\mathsf{max} = 7$, the three announcements cannot be made (without lying). And if $\mathsf{max} = 14$, it is not always the case that Anne knows her number: for example, if Bill has 1 and Cath has 3, Anne cannot determine whether her number is 2 or 4; 213 and 413 are in that case the *only* two triples where Anne is still uncertain. For this sort of finetuning a model checker was helpful and even essential [21]. The interplay between designing riddles, model checking, and dynamic epistemic logical analyses, is a good playing ground between theory and practice of multi-agent system dynamics.

References

1. Alchourrón, C.E., Gärdenfors, P., Makinson, D.: On the logic of theory change. Journal of Symbolic Logic 50, 510–530 (1985)
2. Aucher, G.: A combined system for update logic and belief revision. In: Barley, M.W., Kasabov, N. (eds.) PRIMA 2004. LNCS (LNAI), vol. 3371, pp. 1–17. Springer, Heidelberg (2005)
3. Aumann, R.J.: Agreeing to disagree. Annals of Statistics 4(6), 1236–1239 (1976)
4. Baltag, A., Moss, L.S., Solecki, S.: The logic of public announcements, common knowledge, and private suspicions. In: Gilboa, I. (ed.) TARK 98, pp. 43–56 (1998)
5. Baltag, A., Smets, S.: Dynamic belief revision over multi-agent plausibility models. In: Proceedings of LOFT 2006 (to appear)
6. Fagin, R., Halpern, J.Y., Moses, Y., Vardi, M.Y.: Reasoning about Knowledge. MIT Press, Cambridge, MA (1995)
7. Gerbrandy, J.D.: Bisimulations on Planet Kripke. PhD thesis, University of Amsterdam, ILLC Dissertation Series DS-1999-01 (1999)
8. Hintikka, J.: Knowledge and Belief. Cornell University Press, Ithaca, NY (1962)
9. Kooi, B.P.: Expressivity and completeness for public update logics via reduction axioms. Journal of Applied Non-Classical Logics, 2007 (to appear)

10. Lewis, D.K.: Convention, a Philosophical Study. Harvard University Press, Cambridge (MA) (1969)
11. Liu, A.: Problem section: Problem 182. Math Horizons 11(324) (2004)
12. McCarthy, J.: Formalization of two puzzles involving knowledge. In: Lifschitz, V. (ed.) Formalizing Common Sense: Papers by John McCarthy. Ablex Series in Artificial Intelligence, Ablex Publishing, Greenwich (1990)
13. Meyer, J.-J., Ch., van der Hoek, W.: Epistemic Logic for AI and Computer Science. In: Cambridge Tracts in Theoretical Computer Science 41. Cambridge University Press, Cambridge (1995)
14. Moore, G.E.: A reply to my critics (The Library of Living Philosophers). In: Schilpp, P.A. (ed.) The Philosophy of G.E. Moore, vol. 4, pp. 535–677 (1942)
15. Moses, Y.O., Dolev, D., Halpern, J.Y.: Cheating husbands and other stories: a case study in knowledge, action, and communication. Distributed Computing 1(3), 167–176 (1986)
16. Plaza, J.A.: Logics of public communications. In: Emrich, M., Pfeifer, M., Hadzikadic, M., Ras, Z. (eds.) 4th Int. Symp. on Methodologies for Intelligent Systems, pp. 201–216 (1989)
17. Qian, L.: Sentences true after being announced, `www.stanford.edu/group/nasslli/student/qian.ps`
18. van Benthem, J.F.A.K., van Eijck, J., Kooi, B.P.: Logics of communication and change. Information and Computation 204(11), 1620–1662 (2006)
19. van Ditmarsch, H.P.: Knowledge games. PhD thesis, University of Groningen, ILLC Dissertation Series DS-2000-06 (2000)
20. van Ditmarsch, H.P.: Prolegomena to dynamic logic for belief revision. Synthese (Knowledge, Rationality & Action) 147, 229–275 (2005)
21. van Ditmarsch, H.P., Ruan, J.: Model checking logic puzzles. In: Quatrièmes Journées Francophones MODÈLES FORMELS de l'INTERACTION, Paris, Cahiers du Lamsade, 2007 (to appear)
22. van Ditmarsch, H.P., van der Hoek, W., Kooi, B.P.: Dynamic epistemic logic with assignment. In: AAMAS 05, pp. 141–148. ACM Inc. New York (2005)
23. van Ditmarsch, H.P., van der Hoek, W., Kooi, B.P.: Dynamic Epistemic Logic. In: Synthese Library, vol. 337, Springer, Heidelberg (2007)
24. van Eijck, J.: Dynamic epistemic modelling. Technical report, Centrum voor Wiskunde en Informatica, Amsterdam, (2004) CWI Report SEN-E0424
25. Wu, W.: 100 prisoners and a lightbulb (2001), `http://www.ocf.berkeley.edu/~wwu/papers/100prisonersLightBulb.pdf`