Library of Congress Catalog Data
ISSN: 1095-5054

Dynamic Epistemic Logic

First published Fri Jun 24, 2016

Dynamic Epistemic Logic is the study of modal logics of model change. DEL (pronounced "dell") is a highly active area of applied logic that touches on topics in many areas, including Formal and Social Epistemology, Epistemic and Doxastic Logic, Belief Revision, multi-agent and distributed systems, Artificial Intelligence, Defeasible and Non-monotonic Reasoning, and Epistemic Game Theory. This article surveys DEL, identifying along the way a number of open questions and natural directions for further research.

---

1. Introduction

Dynamic Epistemic Logic is the study of a family of modal logics, each of which is obtained from a given logical language by adding one or more modal operators that describe model-transforming actions. If $[A]$ is such a modality, then new formulas of the form $[A]F$ are used to express the statement that F is true after the occurrence of action A. To determine whether $[A]F$ is true at a pointed Kripke model (M, w) (see Appendix A for definitions), we transform the current Kripke model M according to the prescription of action A and we obtain a new pointed Kripke model (M', w') at which we then investigate whether F is true. If it is true there, then we say that original formula $[A]F$ is true in our starting situation (M, w). If F is not true in the newly produced situation (M', w'), then we conclude the opposite: $[A]F$ is not true in our starting situation (M, w). In this way, we obtain the meaning of $[A]F$ not by the analysis of what obtains in a single Kripke model but by the analysis of what obtains as a result of a specific modality-specified Kripke model transformation. This is a shift from a *static* semantics of truth that takes place in an individual Kripke model to a *dynamic* semantics of truth that takes place across

modality-specified Kripke model transformations. The advantage of the dynamic perspective is that we can analyze the epistemic and doxastic consequences of actions such as public and private announcements without having to “hard wire” the results into the model from the start. Furthermore, we may look at the consequences of different sequences of actions simply by changing the sequence of action-describing modalities.

In the following sections, we will look at the many model-changing actions that have been studied in Dynamic Epistemic Logic. Many natural applications and questions arise as part of this study, and we will see some of the results obtained in this work. Along the way it will be convenient to consider many variations of the general formal setup described above. Despite these differences, at the core is the same basic idea: new modalities describing certain application-specific model-transforming operations are added to an existing logical language and the study proceeds from there. Proceeding now ourselves, we begin with what is perhaps the quintessential and most basic model-transforming operation: the public announcement.

2. Public communication

2.1 Public Announcement Logic

Public Announcement Logic (PAL) is the modal logic study of knowledge, belief, and public communication. PAL (pronounced “pal”) is used to reason about knowledge and belief and the changes brought about in knowledge and belief as per the occurrence of completely trustworthy, truthful announcements. PAL’s most common motivational examples include the *Muddy Children Puzzle* and the *Sum and Product Puzzle* (see, e.g., Plaza 1989, 2007). The *Cheryl’s Birthday* problem, which became a sensation on the Internet in April 2015, can also be addressed using PAL. Here we present a version of the Cheryl’s Birthday problem due to Chang

(2015, 15 April) and a three-child version of the Muddy Children Puzzle (Fagin et al. 1995). Instead of presenting the traditional Sum and Product Puzzle (see Plaza (1989, 2007) for details), we present our own simplification that we call the *Sum and Least Common Multiple Problem*.

Cheryl's Birthday (version of Chang (2015, 15 April)). Albert and Bernard just met Cheryl. "When's your birthday?" Albert asked Cheryl.

Cheryl thought a second and said, "I'm not going to tell you, but I'll give you some clues". She wrote down a list of 10 dates:

- May 15, May 16, May 19
- June 17, June 18
- July 14, July 16
- August 14, August 15, August 17

"My birthday is one of these", she said.

Then Cheryl whispered in Albert's ear the month—and only the month—of her birthday. To Bernard, she whispered the day, and only the day.

"Can you figure it out now?" she asked Albert.

Albert: I don't know when your birthday is, but I know Bernard doesn't know either.
Bernard: I didn't know originally, but now I do.
Albert: Well, now I know too!

When is Cheryl's birthday?

The Muddy Children Puzzle. Three children are playing in the mud. Father calls the children to the house, arranging them in a semicircle

so that each child can clearly see every other child. "At least one of you has mud on your forehead", says Father. The children look around, each examining every other child's forehead. Of course, no child can examine his or her own. Father continues, "If you know whether your forehead is dirty, then step forward now". No child steps forward. Father repeats himself a second time, "If you know whether your forehead is dirty, then step forward now". Some but not all of the children step forward. Father repeats himself a third time, "If you know whether your forehead is dirty, then step forward now". All of the remaining children step forward. How many children have muddy foreheads?

The Sum and Least Common Multiple Puzzle. Referee reminds Mr. S and Mr. L that the least common multiple ("lcm") of two positive integers x and y is the smallest positive integer that is divisible without any remainder by both x and y (e.g., $\mathrm{lcm}(3,6) = 6$ and $\mathrm{lcm}(5,7) = 35$). Referee then says,

> Among the integers ranging from 2 to 7, including 2 and 7 themselves, I will choose two different numbers. I will whisper the sum to Mr. S and the least common multiple to Mr. L.

Referee then does as promised. The following dialogue then takes place:

Mr. S: I know that you don't know the numbers.
Mr. L: Ah, but now I do know them.
Mr. S: And so do I!

What are the numbers?

The Sum and Product Puzzle is like the Sum and Least Common Multiple Puzzle except that the allowable integers are taken in the range $2, \ldots, 100$

(inclusive), Mr. L is told the product of the two numbers (instead of their least common multiple), and the dialogue is altered slightly (L: "I don't know the numbers", S: "I knew you didn't know them", L: "Ah, but now I do know them", S: "And now so do I!"). These changes result in a substantially more difficult problem. See Plaza (1989, 2007) for details.

The reader is advised to try solving the puzzles himself or herself and to read more about PAL below before looking at the PAL-based solutions found in a Appendix B. Later, after the requisite basics of PAL have been presented, the authors will again point the reader to this appendix.

There are many variations of these puzzles, some of which motivate logics that can handle more than just public communication. Restricting attention to the variations above, we note that a formal logic for reasoning about these puzzles must be able to represent various agents' knowledge along with changes in this knowledge that are brought about as a result of public announcements. One important thing to note is that the announcements in the puzzles are all *truthful* and *completely trustworthy*: so that we can solve the puzzles, we tacitly assume (among other things) that everything that is announced is in fact true and that all agents accept the content of a public announcement without question. These assumptions are of course unrealistic in many everyday situations, and, to be sure, there are more sophisticated Dynamic Epistemic Logics that can address more complicated and nuanced attitudes agents may have with respect to the information they receive. Nevertheless, in an appropriately restricted situation, Public Announcement Logic provides a basic framework for reasoning about truthful, completely trustworthy public announcements.

Given a nonempty set $\mathcal{P}$ of propositional letters and a finite nonempty set $\mathcal{A}$ of agents, the basic modal language (ML) is defined as follows:

$$F ::= p \mid F \wedge F \mid \neg F \mid [a]F \qquad p \in \mathcal{P},\ a \in \mathcal{A} \tag{ML}$$

Formulas $[a]F$ are assigned a reading that is doxastic ("agent a believes F") or epistemic ("agent a knows F"), with the particular reading depending on the application one has in mind. In this article we will use both readings interchangeably, choosing whichever is more convenient in a given context. In the language (ML), Boolean connectives other than negation $\neg$ and conjunction $\wedge$ are taken as abbreviations in terms of negation in conjunction as is familiar from any elementary Logic textbook. See Appendix A for further details on (ML) and its Kripke semantics.

The language (PAL) of Public Announcement Logic extends the basic modal language (ML) by adding formulas $[F!]G$ to express that "after the public announcement of F, formula G is true":

$$F ::= p \mid F \wedge F \mid \neg F \mid [a]F \mid [F!]F \qquad p \in \mathcal{P},\ a \in \mathcal{A} \tag{PAL}$$

Semantically, the formula $[F!]G$ is interpreted in a Kripke model as follows: to say that $[F!]G$ is true means that, whenever F is true, G is true after we eliminate all not-F possibilities (and all arrows to and from these possibilities). This makes sense: since the public announcement of F is completely trustworthy, all agents respond by collectively eliminating all non-F possibilities from consideration. So to see what obtains after a public announcement of F occurs, we eliminate the non-F worlds and then see what is true in the resulting situation. Formally, (PAL)-formulas are evaluated as an extension of the binary truth relation $\models$ between pointed Kripke models and (ML)-formulas (defined in Appendix A) as follows: given a Kripke model $M = (W, R, V)$ and a world $w \in W$,

- $M, w \models p$ holds if and only if $w \in V(p)$;
- $M, w \models F \wedge G$ holds if and only if both $M, w \models F$ and $M, w \models G$;
- $M, w \models \neg F$ holds if and only if $M, w \not\models F$;
- $M, w \models [a]F$ holds if and only if $M, v \models F$ for each v satisfying

$wR_a v$; and

- $M, w \models [F!]G$ holds if and only if we have that $M, w \not\models F$ or that $M[F!], w \models G$, where the model

$$M[F!] = (W[F!], R[F!], V[F!])$$

is defined by:

  - $W[F!] := \{v \in W \mid M, v \models F\}$ — retain only the worlds where F is true,
  - $xR[F!]_a y$ if and only if $xR_a y$ — leave arrows between remaining worlds unchanged, and
  - $v \in V[F!](p)$ if an only if $v \in V(p)$ — leave the valuation the same at remaining worlds.

Note that the formula $[F!]G$ is vacuously true if F is false: the announcement of a false formula is inconsistent with our assumption of truthful announcements, and hence every formula follows after a falsehood is announced (*ex falso quodlibet*). It is worth remarking that the dual announcement operator $\langle F! \rangle$ defined by

$$\langle F! \rangle G := \neg [F!] \neg G$$

gives the formula $\langle F! \rangle G$ the following meaning: F is true and, after F is announced, G is also true. In particular, we observe that the announcement formula $\langle F! \rangle G$ is false whenever F is false.

Often one wishes to restrict attention to a class of Kripke models whose relations R_a satisfy certain desirable properties such as reflexivity, transitivity, Euclideanness, or seriality. Reflexivity tells us that agent knowledge is truthful, transitivity tells us that agents know what they know, Euclideanness tells us that agents know what they do not know, and seriality tells us that agent knowledge is consistent. (A belief reading is also possible.) In order to study public announcements over such classes,

we must be certain that the public announcement of a formula F does not transform a given Kripke model M into a new model $M[F!]$ that falls outside of the class. The following theorem indicates when it is that a given class of Kripke models is "closed" under public announcements (meaning a public announcement performed on a model in the class always yields another model in the class).

See Appendix C for the definition of reflexivity, transitivity, Euclideanness, seriality, and other important relational properties.

Public Announcement Closure Theorem. Let $M = (W, R, V)$ be a Kripke model and F be a formula true at at least one world in W.

- If R_a is reflexive, then so is $R[F!]_a$.
- If R_a is transitive, then so is $R[F!]_a$.
- If R_a is Euclidean, then so is $R[F!]_a$.
- If R_a is serial and Euclidean, then so is $R[F \wedge \bigwedge_{x \in \mathcal{A}} \langle x \rangle F!]_a$.

The Public Announcement Closure Theorem tells us that reflexivity, transitivity, and Euclideanness are always closed under the public announcement operation. Seriality is in general not; however, if seriality comes with Euclideanness, then public announcements of formulas of the form $F \wedge \bigwedge_{x \in \mathcal{A}} \langle x \rangle F$ (read, "F is true and consistent with each agent's knowledge") preserve both seriality and Euclideanness. Therefore, if we wish to study classes of models that are serial, then, to make use of the above theorem, we will need to further restrict to models that are both serial and Euclidean and we will need to restrict the language of public announcements so that all announcement formulas have this form. (One could also restrict to another form, so long as public announcements of this form preserve seriality over some class C of serial models.) Restricting the language (PAL) by requiring that public announcements have the form $F \wedge \bigwedge_{x \in \mathcal{A}} \langle x \rangle F$ leads to the language (sPAL) of *serial*

Public Announcement Logic, which we may use when interested in serial and Euclidean Kripke models.

$$F ::= p \mid F \wedge F \mid \neg F \mid [a]F \mid [F \wedge \bigwedge_{x \in \mathcal{A}} \langle x \rangle F!]F$$
$$p \in \mathcal{P},\ a \in \mathcal{A} \qquad \text{(sPAL)}$$

Given a class of Kripke models satisfying certain properties and a modal logic L in the language (ML) that can reason about that class, we would like to construct a Public Announcement Logic whose soundness and completeness are straightforwardly proved. To do this, we would like to know in advance that L is sound and complete with respect to the class of models in question, that some public announcement extension (L+PAL) of the language (ML) (e.g., the language (sPAL) or maybe even (PAL) itself) will include announcements that do not spoil closure, and that there is an easy way for us to determine the truth of (L+PAL)-formulas by looking only the underlying modal language (ML). This way, we can "reduce" completeness of the public announcement theory to the completeness of the basic modal theory L. We call such theories for which this is possible *PAL-friendly*.

PAL-friendly theory. To say that a logic L is *PAL-friendly* means we have the following:

- L is a normal multi-modal logic in the language (ML) (i.e., with modals $[a]$ for each agent $a \in \mathcal{A}$),
- there is a class of Kripke models $\mathcal{C}$ such that L is sound and complete with respect to the collection of pointed Kripke models based on models in $\mathcal{C}$, and
- there is a language (L+PAL) (the "announcement extension of L") obtained from (PAL) by restricting the form of public announcement modals $[F!]$ such that $\mathcal{C}$ is closed under public announcements of this form (i.e., performing a public announcement of this form on a model in $\mathcal{C}$ having at least one world at which the announced formula is true yields another model in $\mathcal{C}$).

See Appendix D for the exact meaning of the first component of a PAL-friendly theory.

Examples of PAL-friendly theories include the common "logic of belief" (multi-modal $\mathsf{KD45}$), the common "logic of knowledge" (multi-modal $\mathsf{S5}$), multi-modal K, multi-modal T, multi-modal $\mathsf{S4}$, and certain logics that mix modal operators of the previously mentioned types (e.g., $\mathsf{S5}$ for $[a]$ and T for all other agent modal operators $[b]$). Fixing a PAL-friendly theory L, we easily obtain an axiomatic theory of public announcement logic based on L as follows.

The axiomatic theory PAL.

- Axiom schemes and rules for the PAL-friendly theory L
- Reduction axioms (all in the language (L+PAL)):
  1. $[F!]p \leftrightarrow (F \to p)$ for letters $p \in \mathcal{P}$
     "After a false announcement, every letter holds—a contradiction. After a true announcement, letters retain their truth values."
  2. $[F!](G \wedge H) \leftrightarrow ([F!]G \wedge [F!]H)$
     "A conjunction is true after an announcement iff each conjunct is."
  3. $[F]\neg G \leftrightarrow (F \to \neg [F!]G)$
     "G is false after an announcement iff the announcement, whenever truthful, does not make G true."
  4. $[F!][a]G \leftrightarrow (F \to [a][F!]G)$
     "a knows G after an announcement iff the announcement, whenever truthful, is known by a to make G true."

- Announcement Necessitation Rule: from G, infer $[F!]G$ whenever the latter is in (L+PAL).
  "A validity holds after any announcement."

The reduction axioms characterize truth of an announcement formula $[F!]G$ in terms of the truth of other announcement formulas $[F!]H$ whose post-announcement formula H is less complex than the original post-announcement formula G. In the case where G is just a propositional letter p, Reduction Axiom 1 says that the truth of $[F!]p$ can be reduced to a formula not containing any announcements of F. So we see that the reduction axioms "reduce" statements of truth of complicated announcements to statements of truth of simpler and simpler announcements until the mention of announcements is not necessary. For example, writing the reduction axiom used in a parenthetical subscript, we have the following sequence of provable equivalences:

$$\begin{aligned} & [[b]p!](p \wedge [a]p) \\ \leftrightarrow_{(2)} \quad & [[b]p!]p \wedge [[b]p!][a]p \\ \leftrightarrow_{(1)} \quad & ([b]p \to p) \wedge [[b]p!][a]p \\ \leftrightarrow_{(4)} \quad & ([b]p \to p) \wedge ([b]p \to [a][[b]p!]p) \\ \leftrightarrow_{(1)} \quad & ([b]p \to p) \wedge ([b]p \to [a]([b]p \to p)) \end{aligned}$$

Notice that the last formula does not contain public announcements. Hence we see that the reduction axioms allow us to express the truth of the announcement-containing formula $[[b]p!](p \wedge [a]p)$ in terms of a provably equivalent announcement-free formula. This is true in general.

PAL **Reduction Theorem.** Given a PAL-friendly theory L, every F in the language (L+PAL) of Public Announcement Logic (without common knowledge) is PAL-provably equivalent to a formula F° coming from the announcement-free fragment of (L+PAL).

The Reduction Theorem makes proving completeness of the axiomatic theory with respect to the appropriate class of pointed Kripke models easy: since every (L+PAL)-formula can all be expressed using a provably equivalent announcement-free (ML)-formula, completeness of the theory PAL follows by the Reduction Theorem, the soundness of PAL, and the known completeness of the underlying modal theory L.

PAL **Soundness and Completeness.** PAL is sound and complete with respect to the collection $\mathcal{C}_*$ of pointed Kripke models for which the underlying PAL-friendly theory L is sound and complete. That is, for each (L+PAL)-formula F, we have that $\mathsf{PAL} \vdash F$ if and only if $\mathcal{C}_* \models F$.

One interesting PAL-derivable scheme (available if allowed by the language (L+PAL)) is the following:

$$[F!][G!]H \leftrightarrow [F \wedge [F!]G!]H$$

This says that two consecutive announcements can be combined into a single announcement: to announce that F is true and then to announce that G is true will have the same result as announcing the single statement that "F is true and, after F is announced, G is true".

We conclude with a few complexity results for Public Announcement Logic.

PAL Complexity. Let $\mathcal{C}$ be the class of all Kripke models. Let $\mathcal{C}_{\mathsf{S5}}$ be the class of Kripke models such that each binary accessibility relation is reflexive, transitive, and symmetric.

- The satisfiability problem for single-agent (PAL) over $\mathcal{C}_{\mathsf{S5}}$ is NP-complete (Lutz 2006).
- The satisfiability problem for multi-agent (PAL) over $\mathcal{C}_{\mathsf{S5}}$ is

PSPACE-complete (Lutz 2006).

- The model checking problem for (PAL) over C is in P (Kooi and van Benthem 2004).

One thing to note about the theory PAL as presented above is that it is parameterized on a PAL-friendly logic L. Therefore, “Public Announcement Logic” as an area of study in fact encompasses a wide-ranging family individual Public Announcement Logics, one for each instance of L. Unless otherwise noted, the results and concepts we present apply to all logics within this family.

In Appendix E, we detail further aspects of Public Announcement Logic: schematic validity, expressivity and succinctness, Gerbrandy–Groeneveld announcements, consistency-preserving announcements and Arrow Update Logic, and quantification over public announcements in Arbitrary Public Announcement Logic.

While iterated public announcements seem like a natural operation to consider (motivated by, e.g., the Muddy Children Puzzle), Miller and Moss (2005) showed that a logic of such a language cannot be recursively axiomatized.

Finally, PAL-based solutions to the Cheryl’s Birthday, Muddy Children, and Sum and Least Common Multiple Puzzles are presented in Appendix B.

2.2 Group knowledge: common knowledge and distributed knowledge

2.2.1 Common knowledge

To reason about common knowledge and public announcements, we add the common knowledge operators $[B*]$ to the language for each group of

agents $B \subseteq \mathcal{A}$. The formula $[B*]F$ is read, ”it is common knowledge among the group B that F is true”. We define the language (PAL+C) of *public announcement logic with common knowledge* as follows:

$$\begin{gathered} F ::= p \mid F \wedge F \mid \neg F \mid [a]F \mid [F!]F \mid [B*]F \\ p \in \mathcal{P},\ a \in \mathcal{A},\ B \subseteq \mathcal{A} \end{gathered} \tag{PAL+C}$$

The semantics of this language over pointed Kripke models is defined in Appendix A. We recall two key defined expressions:

$[B]F$ denotes $\bigwedge_{a\in B}[a]F$ — “everyone in group B knows (or believes) F”;

$[C]F$ denotes $[\mathcal{A}*]F$ — “it is common knowledge (or belief) that F is true.”

For convenience in what follows, we will adopt the epistemic (i.e., knowledge) reading of formulas in the remainder of this subsection. In particular, using the language (PAL+C), we are able to provide a formal sense in which public announcements bring about common knowledge.

Theorem. For each pointed Kripke model (M, w), we have:

- $M, w \models [p!][C]p$ for each propositional letter $p \in \mathcal{P}$.
  “A propositional letter becomes common knowledge after it is announced.”
- If F is successful (i.e., $\models [F!]F$), then $M, w \models [F!][C]F$.
  “A successful formula becomes common knowledge after it is announced.”

We now examine the axiomatic theory of public announcement logic with common knowledge.

The axiomatic theory PAL+C**.**

- Axiom schemes and rules for the theory PAL
- Axiom schemes for common knowledge:
  - $[B*](F \to G) \to ([B*]F \to [B*]G)$
    "Common knowledge is closed under logical consequence."
  - $[B*]F \leftrightarrow (F \wedge [B][B*]F)$, the "Mix axiom"
    "Common knowledge is equivalent to truth and group knowledge of common knowledge."
  - $[B*](F \to [B]F) \to (F \to [B*]F)$, the "Induction axiom"
    "If there is common knowledge that truth implies group knowledge and there is truth, then there is common knowledge."
- CK Necessitation Rule: from F, infer $[B*]F$
  "There is common knowledge of every validity."
- Announcement-CK Rule: from $H \to [F!]G$ and $(H \wedge F) \to [B]H$, infer $H \to [F!][B*]G$
  "If H guarantees the truth of G after F is announced and the joint truth of H and F guarantees group knowledge of H, then H guarantees the announcement of F will lead to common knowledge of G."

$\mathsf{PAL}+C$ Soundness and Completeness (Baltag, Moss, and Solecki 1998, 1999; see also van Ditmarsch, van der Hoek, and Kooi 2007). $\mathsf{PAL}+C$ is sound and complete with respect to the collection $\mathcal{C}_*$ of pointed Kripke models for which the underlying public announcement logic PAL is sound and complete. That is, for each (PAL+C)-formula F, we have that $\mathsf{PAL}+C \vdash F$ if and only if $\mathcal{C}_* \models F$.

Unlike the proof of completeness for the logic PAL without common knowledge, the proof for the logic $\mathsf{PAL}+C$ with common knowledge does not proceed by way of a reduction theorem. This is because adding common knowledge to the language strictly increases the expressivity.

Theorem (Baltag, Moss, and Solecki 1998, 1999; see also van Ditmarsch, van der Hoek, and Kooi 2007). Over the class of all pointed Kripke models, the language (PAL+C) of public announcement logic with common knowledge is strictly more expressive than language (PAL) without common knowledge. In particular, the (PAL+C)-formula $[p!][C]q$ cannot be expressed in (PAL) with respect to the class of all pointed Kripke models: for every (PAL)-formula F there exists a pointed Kripke model (M, w) such that $M, w \not\models F \leftrightarrow [p!][C]q$.

This result rules out the possibility of a reduction theorem for $\mathsf{PAL}+C$: we cannot find a public announcement-free equivalent of every (PAL+C)-formula. This led van Benthem, van Eijck, and Kooi (2006) to develop a common knowledge-like operator for which a reduction theorem does hold. The result is the binary *relativized common knowledge* operator $[B*](F|G)$, which is read, "F is common knowledge among group B relative to the information that G is true". The language (RCK) of relativized common knowledge is given by the following grammar:

$$F ::= p \mid F \wedge F \mid \neg F \mid [a]F \mid [F!]F \mid [B*](F|F) \\ p \in \mathcal{P},\ a \in \mathcal{A},\ B \subseteq \mathcal{A} \qquad \text{(RCK)}$$

and the language (RCK+P) of relativized common knowledge with public announcements is obtained by adding public announcements to (RCK):

$$F ::= p \mid F \wedge F \mid \neg F \mid [a]F \mid [F!]F \mid [B*](F|F) \mid [F!]F \\ p \in \mathcal{P},\ a \in \mathcal{A},\ B \subseteq \mathcal{A} \qquad \text{(RCK-}$$

The semantics of (RCK) is an extension of the semantics of (ML), and the semantics of (RCK+P) is an extension of the semantics of (PAL). In each case, the extension is obtained by adding the following inductive truth clause:

- $M, w \vDash [B*](F|G)$ holds if and only if $M, v \vDash F$ for each v satisfying $w(R[G!]_B)^* v$

Here we recall that $R[G!]$ is the function that obtains after the public announcement of G; that is, we have $xR[G!]_a y$ if and only if x and y are in the model after the announcement of G (i.e., $M, w \vDash G$ and $M, y \vDash G$) and there is an a-arrow from x to y in the original model (i.e., $xR_a y$). The relation $R[G!]_B$ is then the union of the relations for those agents in B; that is, we have $xR[G!]_B y$ if and only if there is an $a \in B$ with $xR[G!]_a y$. Finally, $(R[G!]_B)^*$ is the reflexive-transitive closure of the relation $R[G!]_B$; that is, we have $x(R[G!]_B)^* y$ if and only if $x = y$ or there is a finite sequence

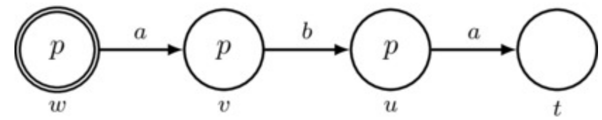
$$x\, R[G!]_B\, z_1\, R[G!]_B \cdots R[G!]_B\, z_n\, R[G!]_B\, y$$

of $R[G!]_B$-arrows connecting x to y. So, all together, the formula $[B*](F|G)$ is true at w if and only if an F-world is at the end of every finite path (of length zero or greater) that begins at w, contains only G-worlds, and uses only arrows for agents in B. Intuitively, this says that if the agents in B commonly assume G is true in jointly entertaining possible alternatives to the given state of affairs w, then, relative to this assumption, F is common knowledge among those in B.

As observed by van Benthem, van Eijck, and Kooi (2006), relativized common knowledge is not the same as non-relativized common knowledge after an announcement. For example, over the collection of all pointed Kripke models, the following formulas are not equivalent:

- $\neg[\{a,b\}*]([a]p \mid p)$ — "it is not the case that, relative to p, it is common knowledge among a and b that a knows p."
- $[p!]\neg[\{a,b\}*][a]p$ — "after p is announced, it is not the case that it is common knowledge among a and b that a knows p."

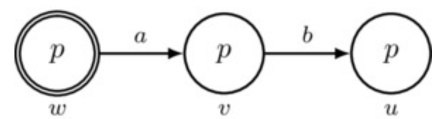
In particular, in the pointed model (M, w) pictured in Figure 1, the formula $\neg[\{a,b\}*]([a]p \mid p)$ is true because there is a path that begins at w, contains only p-worlds, uses only arrows in $\{a,b\}$, and ends on the $\neg[a]p$-world u.



M

Figure 1: The pointed Kripke model (M, w).

However, the formula $[p!]\neg[\{a,b\}*][a]p$ is false at (M, w) because, after the announcement of p, the model $M[p!]$ pictured in Figure 2 obtains, and all worlds in this model are $[a]p$-worlds. In fact, whenever p is true, the formula $[p!]\neg[\{a,b\}*][a]p$ is *always* false: after the announcement of p, all that remains are p-worlds, and therefore every world is an $[a]p$-world.



$M[p!]$

Figure 2: The pointed Kripke model $(M[p!], w)$.

The axiomatic theories of relativized common knowledge with and without public announcements along with expressivity results for the corresponding languages are detailed in Appendix F.

We now state two complexity results for the languages of this subsection.

(PAL+C) **and** (RCK) **Complexity.** Let C be the class of all Kripke models. Let C_{S5} be the class of Kripke models such that each binary accessibility relation is reflexive, transitive, and symmetric.

- The satisfiability problem for each of (PAL+C) and (RCK) over C_{S5} is EXPTIME-complete (Lutz 2006).
- The model checking problem for each of (PAL+C) and (RCK) over C is in P (Kooi and van Benthem 2004).

In the remainder of the article, unless otherwise stated, we will generally assume that we are working with languages that do not contain common knowledge or relativized common knowledge.

2.2.2 Distributed knowledge

Another notion of group knowledge is *distributed knowledge* (Fagin et al. 1995). Intuitively, a group B of agents has distributed knowledge that F is true if and only if, were they to pool together all that they know, they would then know F. As an example, if agents a and b are going to visit a mutual friend, a knows that the friend is at home or at work, and b knows that the friend is at work or at the cafe, then a and b have distributed knowledge that the friend is at work: after they pool together what they know, they will each know the location of the friend. Distributed knowledge and public announcements have been studied by Wáng and Ågotnes (2011). Related to this is the study of whether a notion of group knowledge (such as distributed knowledge) satisfies the property that something known by the group can be established via communication; see Roelofsen (2007) for details.

2.3 Moore sentences

It may seem as though public announcements always "succeed", by which we mean that after something is announced, we are guaranteed that that it is true. After all, this is often the purpose of an announcement: by making the announcement, we wish to inform everyone of its truth. However, it is not hard to come up with announcements that are true when announced

but false afterward; that is, not all announcements are successful. Here are a few everyday examples in plain English.

- Agent a, who is visiting Amsterdam for the first time, steps off the plane in the Amsterdam Airport Schiphol and truthfully says, "a has never made a statement in Amsterdam".
  This is unsuccessful because it is "self-defeating": it rules out various past statements, but it itself is one of those ruled out, so the announcement violates what it says.
- Agent a who does not know it is raining, is told, "It is raining but a does not know it".
  This is an example of a *Moore formula*, which are sentences of the form "p is true but agent a does not know p." In the language (ML), Moore formulas have the form $p \wedge \neg[a]p$. An announcement of a Moore formula is unsuccessful because, after the announcement the agent comes to know the first conjunct p (the statement "it is raining" in the example), which therefore falsifies the second conjunct $\neg[a]p$ (the statement "a does not know it is raining" in the example).

The opposite of unsuccessful formulas are the "successful" ones: these are the formulas that are true after they are announced. Here one should distinguish between "performative announcements" that bring about truth by their very occurrence (e.g., a judge says, "The objection is overruled", which has the effect of making the objection overruled) and "informative announcements" that simply inform their listeners of truth (e.g., our mutual friend says, "I live on 207th Street", which has the effect of informing us of something that is already true). Performative announcements are best addressed in a Dynamic Epistemic Logic setting using *factual changes*, a topic discussed in Appendix G. For now our concern will be with informative announcements.

The phenomena of (un)successfulness of announcements was noted early on by Hintikka (1962) but was not studied in detail until the advent of Dynamic Epistemic Logic. In DEL, the explicit language for public announcements provides for an explicit syntactic definition of (un)successfulness.

> **(Un)successful formula (van Ditmarsch and Kooi 2006; see also Gerbrandy 1999).** Let F be a formula in a language with public announcements.
>
> - To say that F is *successful* means that $\models [F!]F$.
>   "A successful formula is one that is always true after it is announced."
> - To say that F is *unsuccessful* means that $F$is not successful (i.e., $\not\models [F!]F$).
>   "An unsuccessful formula is one that may be false after it is announced."

As we have seen, the Moore formula

$$p \wedge \neg[a]p \tag{MF}$$

is unsuccessful: if (MF) is true, then its announcement eliminates all $\neg p$-worlds, thereby falsifying $\neg[a]p$ (since the truth of $\neg[a]p$ requires the existence of an a-arrow leading to a $\neg p$-world).

An example of a successful formula is a propositional letter p. In particular, after an announcement of p, it is clear that p still holds (since the propositional valuation does not change); that is, $[p!]p$. Moreover, as the reader can easily verify, the formula $[a]p$ is also successful.

In considering (un)successful formulas, a natural question arises: can we provide an syntactic characterization of the formulas that are

(un)successful? That is, is there a way for us know whether a formula is (un)successful simply by looking at its form? Building off of the work of Visser et al. (1994) and Andréka, Németi, and van Benthem (1998), van Ditmarsch and Kooi (2006) provide one characterization of some of the successful (PAL+C)-formulas.

> **Theorem (van Ditmarsch and Kooi 2006).** The *preserved formulas* are formed by the following grammar.
>
> $$F ::= p \mid \neg p \mid F \wedge F \mid F \vee F \mid [a]F \mid [\neg F!]F \mid [B*]F$$
> $$p \in \mathcal{P},\ a \in \mathcal{A},\ B \subseteq \mathcal{A}$$
>
> Every preserved formula is successful.

Using a slightly different notion of successfulness wherein a formula F is said to be *successful* if and only if we have that $M, w \models F \wedge \langle a \rangle F$ implies $M[F!], w \models F$ for each pointed Kripke model (M, w) coming from a given class C, Holliday and Icard (2010) provide a comprehensive analysis of (un)successfulness with respect to the class of single-agent $\mathsf{S5}$ Kripke models and with respect to the class of single-agent $\mathsf{KD45}$ Kripke models. In particular, they provide a syntactic characterization of the successful formulas over these classes of Kripke models. This analysis was extended in part to a multi-agent setting by Saraf and Sourabh (2012). The highly technical details of these works are beyond the scope of the present article.

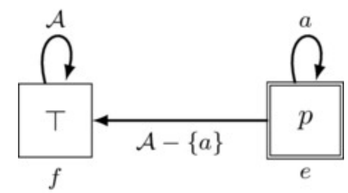
For more on Moore sentences, we refer the reader to Section 5.3 of the *Stanford Encyclopedia of Philosophy* entry on Epistemic Paradoxes (Sorensen 2011).

3. Complex epistemic interactions

In the previous section, we focused on one kind of model-transforming action: the public announcement. In this section, we look at the popular “action model” generalization of public announcements due to Baltag, Moss, and Solecki (Baltag, Moss, and Solecki 1998), together referred to as “BMS”. Action models are simple relational structures that can be used to describe a variety of informational actions, from public announcements to more subtle communications that may contain degrees of privacy, misdirection, deception, and suspicion, to name just a few possibilities.

3.1 Action models describe complex informational scenarios

To begin, let us consider a specific example of a more complex communicative action: a completely private announcement. The idea of this action is that one agent, let us call her a, is to receive a message in complete privacy. Accordingly, no other agent should learn the contents of this message, and, furthermore, no other agent should even consider the possibility that agent a received the message in the first place. (Think of agent a traveling unnoticed to a secret and secure location, finding and reading a coded message only she can decode, and then destroying the message then and there.) One way to think about this action is as follows: there are two possible events that might occur. One of these, let us call it event e, is the announcement that p is true; this is the secret message to a. The other event, let us call it f, is the announcement that the propositional constant $\top$ for truth is true, an action that conveys no new propositional information (since $\top$ is a tautology). Agent a should know that the message is p and hence that the event that is in fact occurring is e. All other agents should mistakenly believe that it is common knowledge that the message is $\top$ and not even consider the possibility that the message is p. Accordingly, other agents should consider event f the one and only possibility and mistakenly believe that this is common knowledge. We picture a diagrammatic representation of this setup in Figure 3.



$\mathsf{Pri}_a(p)$

Figure 3: The pointed action model $(\mathsf{Pri}_a(p), e)$ for the completely private announcement of p to agent a.

In the figure, our two events e and f are pictured as rectangles (to distinguish these from the circled worlds of a Kripke model). The formula appearing inside an event’s rectangle is what is announced when the event occurs. So event e represents the announcement of p, and event f represents the announcement of $\top$. The event that actually occurs, called the “point”, is indicated using a double rectangle; in this case, the point is e. The only event that a considers possible is e because the only a-arrow leaving e loops right back to e. But all of the agents in our agent set $\mathcal{A}$ other than a mistakenly consider the alternative event f as the only possibility: all non–a-arrows leaving e point to f. Furthermore, from the perspective of event f, it is common knowledge that event f (and its announcement of $\top$) is the only event that occurs: every agent has exactly one arrow leaving f and this arrow loops right back to f. Accordingly, the structure pictured above describes the following action: p is to be announced, agent a is to know this, and all other agents are to mistakenly believe it is common knowledge that $\top$ is announced. Structures like those pictured in Figure 3 are called *action models*.

Action model (Baltag, Moss, and Solecki 1998, 1999; see also Baltag and Moss 2004). Other names in the literature: “event model” or “update model”. Given a set of formulas $\mathcal{L}$ and a finite nonempty set $\mathcal{A}$ of agents, an *action model* is a structure

$$A = (E, R, \mathsf{pre})$$

consisting of

- a nonempty finite set E of the possible communicative *events* that might occur,
- a function $R : \mathcal{A} \to P(W \times W)$ that assigns to each agent $a \in \mathcal{A}$ a binary possibility relation $R_a \subseteq E \times E$, and
- a function $\mathsf{pre} : E \to \mathcal{L}$ that assigns to each event $e \in E$ a *precondition* formula $\mathsf{pre}(e) \in \mathcal{L}$. Intuitively, the precondition $\mathsf{pre}(e)$ is announced when event e occurs.

Notation: if A is an action model, then adding a superscript A to a symbol in $\{E, R, \mathsf{pre}\}$ is used to denote a component of the triple that makes up A in such a way that $(E^A, R^A, \mathsf{pre}^A) = A$. We define a *pointed action model*, sometimes also called an *action*, to be a pair (A, e) consisting of an action model A and an event $e \in E^A$ that is called the *point*. In drawing action models, events are drawn as rectangles, and a point (if any) is indicated with a double rectangle. We use many of the same drawing and terminological conventions for action models that we use for (pointed) Kripke models (see Appendix A).

$(\mathsf{Pri}_a(p), e)$ is the action pictured in Figure 3. Given an initial pointed Kripke model (M, w) at which p is true, we determine the model-transforming effect of the action $(\mathsf{Pri}_a(p), e)$ by constructing a new pointed Kripke model

$$(M[\mathsf{Pri}_a(p)], (w, e)).$$

The construction of the Kripke model $M[\mathsf{Pri}_a(p)]$ is given by the BMS "product update".

Product update (Baltag, Moss, and Solecki 1998, 1999; see also Baltag and Moss 2004). Let (M, w) be a pointed Kripke model and (A, e) be a pointed action model. Let $\models$ be a binary satisfaction relation defined between (M, w) and formulas in the language $\mathcal{L}$ of the precondition function $\mathsf{pre}^A : E^A \to \mathcal{L}$ of the action model A. If $M, w \models \mathsf{pre}^A(e)$, then the Kripke model

$$M[A] = (W[A], R[A], V[A])$$

is defined via the *product update* operation $M \mapsto M[A]$ given as follows:

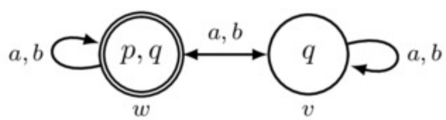
- $W[A] := \{(v,f) \in W \times E \mid M, v \models \mathsf{pre}^A(f)\}$ — pair worlds with events whose preconditions they satisfy,
- $(v_1,f_1)R[A]_a(v_2,f_2)$ if and only if $v_1 R^M_a v_2$ and $f_1 R^A_a f_2$ — insert an a-arrow in $M[A]$ between a pair just in case there is an a-arrow in M between the worlds and an a-arrow in A between the events, and
- $V[A]((v,f)) := V^M(p)$ — make the valuation of p at the pair (v,f) just as it was at v.

An action (A, e) operates on an initial situation (M, w) satisfying $M, w \models \mathsf{pre}^A(e)$ via the product update to produce the resultant situation $(M[A], (w, e))$.

In this definition, the worlds of $M[A]$ are obtained by making multiple copies of the worlds of M, one copy per event $f \in E^A$. The event-f copy of a world v in M is represented by the pair (v,f). Such a pair is to be included in the worlds of $M[A]$ if and only if (M, v) satisfies the precondition $\mathsf{pre}^A(f)$ of event f. The term "product update" comes from the fact that the set $W[A]$ of worlds of $M[A]$ is specified by restricting the full Cartesian product $W^M \times E^A$ to those pairs (v,f) whose indicated world v satisfies the precondition $\mathsf{pre}^A(f)$ of the indicated event f; that is, the "product update" is based on a restricted Cartesian product, hence the name.

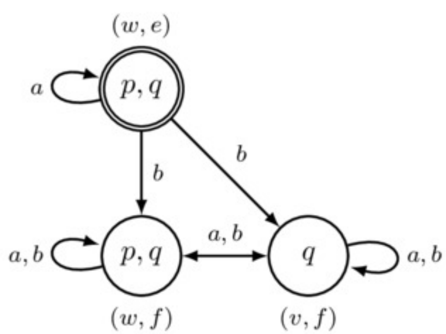
According to the product update, we insert an a-arrow $(v_1,f_1) \to_a (v_2,f_2)$ in $M[A]$ if and only if there is an a-arrow $v_1 \to_a v_2$ in M and an a-arrow $f_1 \to_a f_2$ in A. In this way, agent a's uncertainty in the resultant model $M[A]$ comes from two sources: her initial uncertainty in M (represented by R_a^M) as to which is the actual world and her uncertainty in A (represented by R_a^A) as to which is the actual event. Finally, the valuation at the copy (v,f) in $M[A]$ is just the same as it was at the original world v in M.

For an example of the product update in action, consider the following pointed Kripke model (M,w):



M

The action model $\mathsf{Pri}_a(p)$ from Figure 3 operates on (M,w) via the product update to produce the resultant situation $(M[\mathsf{Pri}_a(p)],(w,e))$ pictured as follows:



$M[\mathsf{Pri}_a(p)]$

Indeed, to produce $M[\mathsf{Pri}_a(p)]$ from M via the product update with the action model $\mathsf{Pri}_a(p)$:

- Event e has us copy worlds at which $\mathsf{pre}^{\mathsf{Pri}_a(p)}(e) = p$ is true; this is just the world w, which we retain in the form (w,e) with the same valuation.
- Event f has us copy worlds at which $\mathsf{pre}^{\mathsf{Pri}_a(p)}(f) = \top$ is true; this is both w and v, which we retain in the forms (w,f) and (v,f), respectively, with their same respective valuations.
- We interconnect the worlds in $M[\mathsf{Pri}_a(p)]$ with agent arrows according to the recipe of the product update: place an arrow between pairs just in case we have arrows componentwise in M and in $\mathsf{Pri}_a(p)$, respectively. For example, we have a b-arrow $(w,e) \to_b (v,f)$ in $M[\mathsf{Pri}_a(p)]$ because we have the b-arrow $w \to_b v$ in M and the b-arrow $e \to_b f$ in $\mathsf{Pri}_a(p)$.
- The point (i.e., actual world) (w,e) of the resultant situation is obtained by paring together the point w from the initial situation (M,w) and the point e from the applied action $(\mathsf{Pri}_a(p),e)$.

We therefore obtain the model $M[\mathsf{Pri}_a(p)]$ as pictured above. We note that the product update-induced mapping

$$(M,w) \mapsto (M[A],(w,e))$$

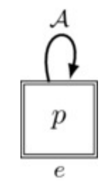
from the initial situation (M,w) to the resultant situation $(M[A],(w,e))$ has the following effect: we go from an initial situation (M,w) in which neither agent knows whether p is true to a resultant situation $(M[A],(w,e))$ in which a knows p is true but b mistakenly believes everyone's knowledge is unchanged. This is of course just what we want of the private announcement of p to agent a.

We now take a moment to comment on the similarities and differences between action models and Kripke models. To begin, both are labeled directed graphs (consisting of labeled nodes and labeled edges pointing between the nodes). A node of a Kripke model (a "world") is labeled by the propositional letters that are true at the world; in contrast, a node of an action model (an "event") is labeled by a single formula that is to be announced if the event occurs. However, in both cases, agent uncertainty is represented using the same "considered possibilities" approach. In the case of Kripke models, an agent considers various possibilities for the world that might be actual; in the case of action models, an agent considers various possibilities for the event that might actually occur. The key insight behind action models, as put forward by Baltag, Moss, and Solecki (1998), is that these two uncertainties can be represented using similar graph-theoretic structures. We can therefore leverage our experience working with Kripke models when we need to devise new action models that describe complex communicative actions. In particular, to construct an action model for a given action, all we must do is break up the action into a number of simple announcement events and then describe the agents' respective uncertainties among these events in the appropriate way so as to obtain the desired action. The difficulty, of course, is in determining the exact uncertainty relationships. However, this determination amounts to inserting the appropriate agent arrows between possible events, and doing this requires the same kind of reasoning as that which we used in the construction of Kripke models meeting certain basic or higher-order knowledge constraints. We demonstrate this now by way of example, constructing a few important action models along the way.

3.2 Examples of action models

We saw the example of a completely private announcement in Figure 3, a complex action in which one agent learns something without the other agents even suspecting that this is so. Before devising an action model for

another similarly complicated action, let us return to our most basic action: the public announcement of p. The idea of this action is that all agents receive the information that p is true, and this is common knowledge. So to construct an action model for this action, we need only one event e that conveys the announcement that p is true, and the occurrence of this event should be common knowledge. This leads us immediately to the action model $\mathsf{Pub}(p)$ pictured in Figure 4.

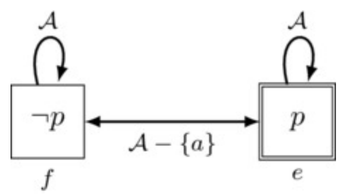


$\mathsf{Pub}(p)$

Figure 4: The pointed action model $(\mathsf{Pub}(p), e)$ for the public announcement of p.

It is not difficult to see that $\mathsf{Pub}(p)$ is just what we want: event e conveys the desired announcement and the reflexive arrows for each agent make it so that this event is common knowledge. It is important to note that in virtue of the fact that we can construct an action model for public announcements, it follows that action models are a generalization of public announcements.

We now turn to a more complicated action: the semi-private announcement of p to agent a (sometimes called the "semi-public announcement" of p to agent a). The idea of this action is that agent a is told that p is true, the other agents know that a is told the truth value of p, but these other agents do not know what it is exactly that a is told. This suggests an action model with two events, one for each thing that a might be told: an event e that announces p and event f that announces $\neg p$. Agent a is to know which event occurs, whereas all other agents are to be uncertain as to which event occurs. This leads us to the action model $\frac{1}{2}\mathsf{Pri}_a(p)$ pictured in Figure 5.



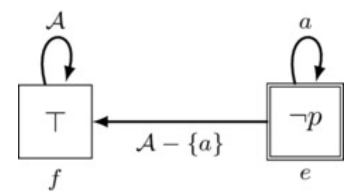
$\frac{1}{2}\mathsf{Pri}_a(p)$

Figure 5: The pointed action model $(\frac{1}{2}\mathsf{Pri}_a(p), e)$ for the semi-private announcement of p to agent a.

We see that $\frac{1}{2}\mathsf{Pri}_a(p)$ satisfies just what we want: the actual event that occurs is the point e (the announcement of the precondition p), agent a knows this, but all other agents consider it possible that either e (the announcement of p) or f (the announcement of $\neg p$) occurred. Furthermore, the other agents know that a knows which event was the case (since at each of the events e and f that they consider possible, agent a knows the event that occurs). This is just what we want of a semi-private announcement.

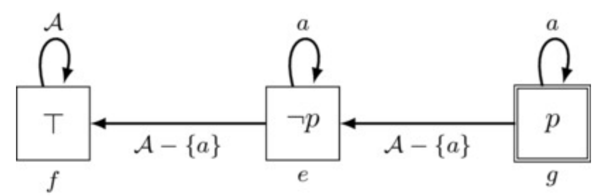
Finally, let us consider a much more challenging action: the misleading private announcement of p to agent a. The idea of this action is that agent a is told p in a completely private manner but all other agents are misled into believing that a received the private announcement of $\neg p$ instead. So to construct an action model for this, we need a few elements: events for the private announcement of $\neg p$ to a that the non-a agents mistakenly believe occurs and an event for the actual announcement of p that only a knows occurs. As for the events for the private announcement of $\neg p$, it follows by a simple modification of Figure 3 that the private

announcement of $\neg p$ to agent a is the action $(\mathsf{Pri}_a(\neg p), e)$ pictured as follows:



$\mathsf{Pri}_a(\neg p)$

Since the other agents are to believe that the above action occurs, they should believe it is event e that occurs. However, they are mistaken: what actually does occur is a new event g that conveys to a the private information that p is true. Taken together, we obtain the action $(\mathsf{MPri}_a(p), g)$ pictured in Figure 6.



$\mathsf{MPri}_a(p)$

Figure 6: The pointed action model $(\mathsf{MPri}_a(p), g)$ for the misleading private announcement of p to agent a.

Looking at $\mathsf{MPri}_a(p)$, we see that if we were to to delete event g (and all arrows to and from g), then we would obtain $\mathsf{Pri}_a(\neg p)$. So events e and f in $\mathsf{MPri}_a(p)$ play the role of representing the "misdirection" the non-a agents experience: the private announcement of $\neg p$ to agent a. However, it is event g that actually occurs: this event conveys to a that p is true while misleading the other agents into believing that it is event e, the event corresponding to the private announcement of $\neg p$ to a, that occurs. In sum, a receives the information that p is true while the other agents are mislead into believing that a received the private announcement of $\neg p$. One consequence of this is that non-a agents come to hold the following beliefs: $\neg p$ is true, agent a knows this, and agent a believes the others believe that no new propositional information was provided. These beliefs are all incorrect. The non-a agents are therefore highly mislead.

3.3 The Logic of Epistemic Actions

Now that we have seen a number of action models, we turn to the formal syntax and semantics of the language (EAL) of *Epistemic Action Logic* (a.k.a., *the Logic of Epistemic Actions*). We define the language (EAL) along with the set AM_* of pointed action models with preconditions in the language (EAL) according to the following recursive grammar:

$$\begin{gathered} F ::= p \mid F \wedge F \mid \neg F \mid [a]F \mid [A,e]F \\ p \in \mathcal{P},\ a \in \mathcal{A},\ (A,e) \in \mathsf{AM}_* \end{gathered} \tag{EAL}$$

To be clear: in the language (EAL), the precondition $\mathsf{pre}^A(e)$ of an action model A may be a formula that includes an action model modality $[A',e']$ for some other action $(A',e') \in \mathsf{AM}_*$. For full technical details on how this works, please see Appendix H.

For convenience, we let AM denote the set of all action models whose preconditions are all in the language (EAL). As we saw in the previous two subsections, the set AM_* contains pointed action models for public announcements (Figure 4), private announcements (Figure 3), semi-private announcements (Figure 5), and misleading private announcements (Figure 6), along with many others. The satisfaction relation $\models$ between pointed Kripke models and formulas of (EAL) is the smallest extension of the relation $\models$ for (ML) (see Appendix A) satisfying the following:

- $M,w \models [A,e]G$ holds if and only if $M,w \not\models \mathsf{pre}^A(e)$ or $M[A],(w,e) \models G$, where the Kripke model $M[A]$ is defined via the BMS product update (Baltag, Moss, and Solecki 1999).

Note that the formula $[A,e]G$ is vacuously true if the precondition $\mathsf{pre}(e)$ of event e is false. Accordingly, the action model semantics retains the assumption of truthfulness that we had for public announcements. That is, for an event to actually occur, its precondition must be true. As a consequence, the occurrence of an event e implies that its precondition $\mathsf{pre}(e)$ was true, and hence the occurrence of an event conveys its precondition formula as a message. If an event can occur at a given world, then we say that the event is *executable* at that world.

> **Executable events and action models.** To say that a pointed action model (A,e) is *executable* at a pointed Kripke model (M,w) means that $M,w \models \mathsf{pre}(e)$. To say that an event f in an action model A is *executable* means that (A,f) is executable. To say that an action model A is *executable* in a Kripke model M means there is an event f in A and a world v in M such that f is executable at (M,v).

As was the case for PAL, one often wishes to restrict attention to Kripke models whose relations R_a satisfy certain desirable properties such as reflexivity, transitivity, Euclideanness, and seriality. In order to study actions over such classes, we must be certain that the actions do not transform a Kripke model in the class into a new Kripke model not in the class; that is, we must ensure that the class of Kripke models is "closed" under actions. The following theorem provides some sufficient conditions that guarantee closure.

> **Action Model Closure Theorem.** Let $M = (W^M, R^M, V)$ be a Kripke model and $A = (W^A, R^A, \mathsf{pre})$ be an action model executable in M.
>
> - If R^M_a and R^A_a are reflexive, then so is $R^M[A]_a$.

- If R_a^M and R_a^A are transitive, then so is $R^M[A]_a$.
- If R_a^M and R_a^A are Euclidean, then so is $R^M[A]_a$.
- If A satisfies the condition that every event $e \in W^A$ gives rise to a nonempty set

$$S(e) \subseteq \{f \in W^A \mid eR_a^A f\}$$

of events such that

$$\models \mathsf{pre}^A(e) \to \langle a \rangle \left(\bigvee_{f \in S(e)} \mathsf{pre}^A(f) \right) ,$$

then $R^M[A]_a$ is serial. (Note: the condition on A and the executability of A in M together imply that R_a^M is serial.)

This theorem, like the analogous theorem for Public Announcement Logic, is used in providing simple sound and complete theories for the Logic of Epistemic Actions based on appropriate "action-friendly" logics.

Action-friendly logic. To say that a logic L is *action-friendly* means we have the following:

- L is a normal multi-modal logic in the language (ML) (i.e., with modals $[a]$ for each agent $a \in \mathcal{A}$),
- there is a class of Kripke models $\mathcal{C}$ such that L is sound and complete with respect to the collection of pointed Kripke models based on models in $\mathcal{C}$, and
- there is a language (L+EAL) (the "action model extension of L") obtained from (EAL) by restricting the form of action models such that $\mathcal{C}$ is closed under the product update with executable actions of this form (i.e., performing an executable action model of this form on a model in $\mathcal{C}$ yields another model in $\mathcal{C}$).

The various axiomatic theories of modal logic with action models (without common knowledge) are obtained based on the choice of an underlying action-friendly logic L.

The axiomatic theory EAL. Other names in the literature: DEL or AM (for "action model"; see van Ditmarsch, van der Hoek, and Kooi 2007).

- Axiom schemes and rules for the action-friendly logic L
- Reduction axioms (each in the language (L+EAL)):
  1. $[A,e]p \leftrightarrow (\mathsf{pre}(e) \to p)$ for letters $p \in \mathcal{P}$
     "After a non-executable action, every letter holds—a contradiction. After an executable action, letters retain their truth values."
  2. $[A,e](G \land H) \leftrightarrow ([A,e]G \land [A,e]H)$
     "A conjunction is true after an action iff each conjunct is."
  3. $[A,e]\neg G \leftrightarrow (\mathsf{pre}(e) \to \neg[A,e]G)$
     "G is false after an action iff the action, whenever executable, does not make G true."
  4. $[A,e][a]G \leftrightarrow (\mathsf{pre}(e) \to \bigwedge_{eR_af}[a][A,f]G)$
     "a knows G after an action iff the action, whenever executable, is known by a to make G true despite her uncertainty of the actual event."
- Action Necessitation Rule: from G, infer $[A,e]G$ whenever the latter is in (L+EAL).
  "A validity holds after any action."

The first three reduction axioms are nearly identical to the corresponding reduction axioms for PAL, except that the first and third EAL reduction axioms check the truth of a precondition in the place where the PAL reduction axioms would check the truth of the formula to be announced. This is actually the same kind of check: for an event, the precondition must hold in order for the event to be executable; for a public announcement, the formula must be true in order for the public announcement to occur (and hence for the public announcement event in question to be "executable"). The major difference between the PAL and EAL reduction axioms is in the fourth EAL reduction axiom. This axiom specifies the conditions under which an agent has belief (or knowledge) of something after the occurrence of an action. In particular, adopting a doxastic reading for this discussion, the axiom says that agent a believes G after the occurrence of action (A, e) if and only if the formula

$$\mathsf{pre}(e) \to \bigwedge_{eR_af}[a][A,f]G$$

is true. This formula, in turn, says that if the precondition is true—and therefore the action is executable—then, for each of the possible events the agent entertains, she believes that G is true if the event in question occurs. This makes sense: a cannot be sure which of the events has occurred, and so for her to believe something after the action has occurred, she must be sure that this something is true no matter which of her entertained events might have been the actual one. For example, if a sees her friend b become elated as he listens to something he hears on the other side of a private phone call, then the a may not know exactly what it is that b is being told; nevertheless, a has reason to believe that b is receiving good news because, no matter what it is exactly that he is being told (i.e., no matter which of the events she thinks that he may be hearing), she knows from his reaction that he must be receiving good news.

As was the case for PAL, the EAL reduction axioms allow us to "reduce" each formula containing action models to a provably equivalent formula whose action model modalities appear before formulas of lesser complexity, allowing us to eliminate action model modalities completely via a sequence of provable equivalences. As a consequence, we have the following.

EAL **Reduction Theorem (Baltag, Moss, and Solecki 1998, 1999; see also Baltag and Moss 2004).** Given an action-friendly logic L, every F in the language (L+EAL) of Epistemic Action Logic (without common knowledge) is EAL-provably equivalent to a formula F° coming from the action model-free modal language (ML).

Once we have proved EAL is sound, the Reduction Theorem leads us to axiomatic completeness via the known completeness of the underlying modal theory.

EAL **Soundness and Completeness (Baltag, Moss, and Solecki 1998, 1999; see also Baltag and Moss 2004).** EAL is sound and complete with respect to the collection $\mathcal{C}_*$ of pointed Kripke models for which the underlying action-friendly logic L is sound and complete. That is, for each (L+EAL)-formula F, we have that $\mathsf{EAL} \vdash F$ if and only if $\mathcal{C}_* \models F$.

We saw above that for PAL it was possible to combine two consecutive announcements into a single announcement via the schematic validity

$$[F!][G!]H \leftrightarrow [F \wedge [F!]G!]H.$$

Something similar is available for action models.

Action model composition. The *composition* $A \circ B = (E, R, \mathsf{pre})$ of action models $A = (E^A, R^A, \mathsf{pre}^A)$ and $B = (E^B, R^B, \mathsf{pre}^B)$ is defined as follows:

- $E = E^A \times E^B$ — composed events are pairs (e,f) of constituent events;
- $(e_1,f_1)R_a(e_2,f_2)$ if and only if $e_1 R^A_a e_2$ and $f_1 R^B_a f_2$ — a composed event is entertained iff its constituent events are; and
- $\mathsf{pre}((e_1, e_2)) = \mathsf{pre}^A(e_1) \wedge [A, e_1]\mathsf{pre}^B(e_2)$ — a composed event

is executable iff the first constituent is executable and, after it occurs, the second constituent is executable as well.

Composition Theorem. Each instance of the following schemes is EAL-derivable (so long as they are permitted in the language (L+EAL)).

- Composition Scheme: $[A,e][B,f]G \leftrightarrow [A \circ B,(e,f)]G$
- Associativity Scheme:
  $[A \circ B,(e,f)][C,g]H \leftrightarrow [A,e][B \circ C,(f,g)]H$

We conclude this subsection with two complexity results for (EAL).

EAL Complexity (Aucher and Schwarzentruber 2013). Let $\mathcal{C}$ be the class of all Kripke models.

- The satisfiability problem for (EAL) over $\mathcal{C}$ is NEXPTIME-complete.
- The model checking problem for (EAL) over $\mathcal{C}$ is PSPACE-complete.

Appendix G provides information on action model equivalence (including the notions of action model bisimulation and emulation), studies a simple modification that enables action models to change the truth value of propositional letters (permitting so-called "factual changes"), and shows how to add common knowledge to EAL.

3.4 Variants and generalizations

In this section, we mention some variants of the action model approach to Kripke model transformation.

- **Graph modifier logics.** Aucher et al. (2009) study extensions of (ML) that contain modalities for performing certain graph modifying

  operations.
- **Generalized Arrow Update Logic.** Kooi and Renne (2011b) introduce a theory of model-changing operations that delete arrows instead of worlds. This theory, which is equivalent to EAL in terms of language expressivity and update expressivity, is a generalization of a simpler theory called Arrow Update Logic (see Section 4 of Appendix E).
- **Logic of Communication and Change.** Van Benthem, van Eijck, and Kooi (2006) introduce LCC, the Logic of Communication and Change, as a Propositional Dynamic Logic-like language that incorporates action models with "factual change".
- **General Dynamic Dynamic Logic.** Girard, Seligman, and Liu (2012) propose General Dynamic Dynamic Logic GDDL, a Propositional Dynamic Logic-style language that has complex action model-like modalities that themselves contain Propositional Dynamic Logic-style instructions.

More on these variants to the action model approach may be found in Appendix I.

4. Belief change and Dynamic Epistemic Logic

Up to this point, the logics we have developed all have one key limitation: an agent cannot meaningfully assimilate information that contradicts her knowledge or beliefs; that is, incoming information that is *inconsistent* with an agent's knowledge or belief leads to difficulties. For example, if agent a believes p, then announcing that p is false brings about a state in which the agent's beliefs are trivialized (in the sense that she comes to believe *every* sentence):

$$\models [a]p \to [\neg p!][a]F \quad \text{for all formulas } F.$$

Note that in the above, we may replace F by a contradiction such as the propositional constant $\bot$ for falsehood. Accordingly, an agent who initially believes p is lead by an announcement that p is false to an inconsistent state in which she believes *everything*, including falsehoods. This trivialization occurs whenever something is announced that contradicts the agent's beliefs; in particular, it occurs if a contradiction such as $\bot$ is itself announced:

$$\vDash [\bot!][a]F \quad \text{for all formulas } F.$$

In everyday life, the announcement of a contradiction, when recognized as such, is generally not informative; at best, a listener who realizes she is hearing a contradiction learns that there is some problem with the announcer or the announced information itself. However, the announcement of something that is not intrinsically contradictory but merely contradicts existing beliefs is an everyday occurrence of great importance: upon receipt of trustworthy information that our belief about something is wrong, a rational response is to adjust our beliefs in an appropriate way. Part of this adjustment requires a determination of our attitude toward the general reliability or trustworthiness of the incoming information: perhaps we trust it completely, like a young child trusts her parents. Or maybe our attitude is more nuanced: we are willing to trust the information for now, but we still allow for the possibility that it might be wrong, perhaps leading us to later revise our beliefs if and when we learn that it is incorrect. Or maybe we are much more skeptical: we distrust the information for now, but we do not completely disregard the possibility, however seemingly remote, that it might turn out to be true.

What is needed is an adaptation of the above-developed frameworks that can handle incoming information that may contradict existing beliefs and that does so in a way that accounts for the many nuanced attitudes an agent may have with respect to the general reliability or trustworthiness of

the information. This has been a focus of much recent activity in the DEL literature.

4.1 Belief Revision: error-aware belief change

Belief Revision is the study of belief change brought about by the acceptance of incoming information that may contradict initial beliefs (Gärdenfors 2003; Ove Hansson 2012; Peppas 2008). The seminal work in this area is due to Alchourrón, Gärdenfors, and Mackinson, or "AGM" (1985). The AGM approach to belief revision characterizes belief change using a number of postulates. Each postulate provides a qualitative account of the belief revision process by saying what must obtain with respect to the agent's beliefs after revision by an incoming formula F. For example, the AGM *Success* postulate says that the formulas the agent believes after revision by F must include F itself; that is, the revision always "succeeds" in causing the agent to come to believe the incoming information F.

Belief Revision has traditionally restricted attention to single-agent, "ontic" belief change: the beliefs in question all belong to a single agent, and the beliefs themselves concern only the "facts" of the world and not, in particular, higher-order beliefs (i.e., beliefs about beliefs). Further, as a result of the Success postulate, the incoming formula F that brings about the belief change is assumed to be *completely trustworthy*: the agent accepts without question the incoming information F and incorporates it into her set of beliefs as per the belief change process.

Work on belief change in Dynamic Epistemic Logic incorporates key ideas from Belief Revision Theory but removes three key restrictions. First, belief change in DEL can can involve higher-order beliefs (and not just "ontic" information). Second, DEL can be used in multi-agent scenarios. Third, the DEL approach permits agents to have more nuanced attitudes with respect to the incoming information.

4.2 Static and dynamic belief change

The literature on belief change in Dynamic Epistemic Logic makes an important distinction between "static" and "dynamic" belief change (van Ditmarsch 2005; Baltag and Smets 2008b; van Benthem 2007).

- **Static belief change:** the objects of agent belief are fixed external truths that do not change, though the agent's beliefs about these truths may change. In a motto, static belief change involves "changing beliefs about an *unchanging* situation".
- **Dynamic belief change:** the objects of agent belief include not only external truths but also the beliefs themselves, and part or all of these can change. In a motto, dynamic belief change involves "changing beliefs about a *changing* situation that itself includes these very beliefs".

To better explain and illustrate the difference, let us consider the result of a belief change brought about by the Moore formula

$$p \wedge \neg[a]p, \tag{MF}$$

informally read, "p is true but agent a does not believe it". Let us suppose that this formula is true; that is, p is true and, indeed, agent a does not believe that p is true. Now suppose that agent a receives the formula (MF) from a completely trustworthy source and is supposed to change her beliefs to take into account the information this formula provides. In a dynamic belief change, she will accept the formula (MF) and hence, in particular, she will come to believe that p is true. But then the formula (MF) becomes false: she now believes p and therefore the formula $\neg[a]p$ ("agent a does not believe p") is false. So we see that this belief change is

indeed dynamic: in revising her beliefs based on the incoming true formula (MF), the truth of the formula (MF) was itself changed. That is, the "situation", which involves the truth of p and the agent's beliefs about this truth, changed as per the belief change brought about by the agent learning that (MF) is true. (As an aside, this example shows that for *dynamic* belief change, the AGM Success postulate is violated and so must be dropped.)

Perhaps surprisingly, it is also possible to undergo a *static* belief change upon receipt of the true formula (MF) from a completely trustworthy source. For this to happen, we must think of the "situation" with regard to the truth of p and the agent's beliefs about this truth as completely static, like a "snapshot in time". We then look at how the agent's beliefs about that static snapshot might change upon receipt of the completely trustworthy information that (MF) was true in the moment of that snapshot. To make sense of this, it might be helpful to think of it this way: the agent learns something *in the present* about what was true of her situation *in the past*. So her present views about her past beliefs change, but the past beliefs remain fixed. It is as though the agent studies a photograph of herself from the past: her "present self" changes her beliefs about that "past self" pictured in the photograph, fixed forever in time. In a certain respect, the "past self" might as well be a different person:

> Now that I have been told (MF) is true at the moment pictured in the photograph, what can I say about the situation in the picture and about the person in that situation?

So to perform a *static belief change* upon receipt of the incoming formula F, the agent is to change her present belief based on the information that F was true in the state of affairs that existed *before* she was told about F. Accordingly, in performing a static belief change upon receipt of (MF), the agent will come to accept that, just before she was told (MF), the letter

p was true but she did not believe that p was true. But most importantly, this will not cause her to believe that (MF) is true *afterward*: she is only changing her beliefs about what was true *in the past*; she has not been provided with information that bears on the present. In particular, while she will change her belief about the truth of p in the moment that existed just before she was informed of (MF), she will leave her present belief about p as it is (i.e., she still will not know that p is true). Therefore, upon static belief revision by (MF), it is still the case that (MF) is true! (As an aside, this shows that for *static* belief change, the AGM Success postulate is satisfied.)

Static belief change occurs in everyday life when we receive information about something that can quickly change, so that the information can become "stale" (i.e., incorrect) just after we receive it. This happens, for example, with our knowledge of the price of a high-volume, high-volatility stock during trading hours: if we check the price and then look away for the rest of the day, we only know the price at the given moment in the past and cannot guarantee that the price remains the same, even right after we checked it. Therefore, we only know the price of the stock in the past—not in the present—even though for practical reasons we sometimes operate under the fiction that the price remains constant after we checked it and therefore speak as though we know it (even though we really do not).

Dynamic belief change is more common in everyday life. It happens whenever we receive information whose truth cannot rapidly become "stale": we are given the information and this information bears directly on our present situation.

We note that the distinction between static and dynamic belief change may raise a dilemma that bears on the problem of skepticism in Epistemology (see, e.g., entry on Epistemology): our "dynamic belief change skeptic"

might claim that *all* belief changes must be static because we cannot really know that then information we have received has not become stale. To the authors' knowledge, this topic has not yet been explored.

4.3 Plausibility models and belief change

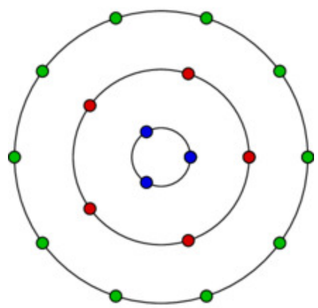
In the DEL study of belief change, situations involving the beliefs of multiple agents are represented using a variation of basic Kripke models called *plausibility models*. Static belief change is interpreted as conditionalization in these models: without changing the model (i.e., the situation), we see what the agent would believe conditional on the incoming information. This will be explained in detail in a moment. Dynamic belief change involves transforming plausibility models: after introducing plausibility model-compatible action models, we use model operators defined from these "plausibility action models" to describe changes in the plausibility model (i.e., the situation) itself.

Our presentation of the DEL approach to belief change will follow Baltag and Smets (2008b), so all theorems and definitions in the remainder of Section 4 are due to them unless otherwise noted. Their work is closely linked with the work of van Benthem (2007), Board (2004), Grove (1988), and others. For an alternative approach based on Propositional Dynamic Logic, we refer the reader to van Eijck and Wang (2008).

Plausibility models are used to represent more nuanced versions of knowledge and belief. These models are also used to reason about *static* belief changes. The idea behind plausibility models is similar to that for our basic Kripke models: each agent considers various worlds as possible candidates for the actual one. However, there is a key difference: among any two worlds w and v that an agent a considers possible, she imposes a relative *plausibility order*. The plausibility order for agent a is denoted by $\geq_a$. We write

$w \geq_a v$ to mean that "world w is no more plausible than world v according to agent a".

Note that if we think of $\geq_a$ as a "greater than or equal to" sign, it is the "smaller" world that is either *more* plausible or else of equal plausibility. The reason for ordering things in this way comes from an idea due to Grove (1988): we think of each world as positioned on the surface of exactly one of a series of concentric spheres (of non-equal radii), with a more plausible world located on a sphere of smaller radius and a less plausible world located on a sphere of greater radius. Consider the following illustration:

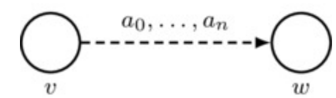


In this diagram, the black concentric circles indicate spheres, the blue points on the smallest (i.e., innermost) sphere are the most plausible worlds overall, the red points on the second-smallest (i.e., middle) sphere are the second-most plausible worlds, and the green points on the largest sphere are the least plausible worlds overall.

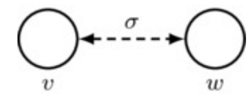
We write $\leq_a$ ("no less plausible than") for the *converse plausibility relation*: $w \leq_a v$ means that $v \geq_a w$. Also, we define the *strict plausibility relation* $>_a$ ("strictly more plausible than") in the usual way: $w >_a v$ means that we have $w \geq_a v$ and $v \not\geq_a w$. (A slash through the relation means the relation does not hold.) The *strict converse plausibility relation* $<_a$ ("strictly less plausible than") is defined as expected: $w <_a v$ means that $v >_a w$. Finally, we define the *equi-plausibility relation* $\simeq_a$ ("equally plausible") as follows: $w \simeq_a v$ means that we have $w \geq_a v$ and $v \geq_a w$.

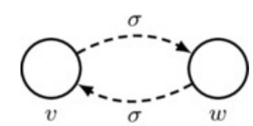
We draw plausibility models much like our basic Kripke models from before except that we use dashed arrows (instead of solid ones) in order to indicate the plausibility relations and also to indicate that the picture in question is one of a plausibility model. We adopt the following conventions for drawing plausibility models.

- One-way arrows indicate non-increasing plausibility:

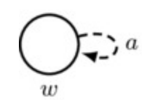


  indicates $v \geq_a w$ for each $a \in \{a_0, \ldots, a_n\}$
- Two-way arrows are a shorthand for two one-way arrows, one in each direction: letting σ denote a comma-separated list of agents in $\mathcal{A}$,

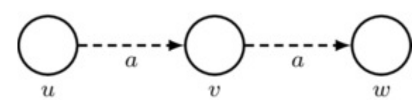


indicates


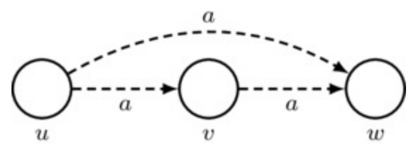
- Reflexive arrows are implied (so that $\geq_a$ is reflexive for each $a \in \mathcal{A}$):

indicates


- Transitive arrows are implied (so that $\geq_a$ is transitive for each $a \in \mathcal{A}$):

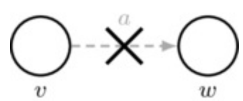


indicates



The transitive arrow rule and the two-way arrow rule may interact: if one or both of the arrows $u \dashrightarrow_a v$ or $v \dashrightarrow_a w$ were two-way, then we would still obtain the implied transitive arrow $u \dashrightarrow_a w$.

- An absence of a drawn or implied a-arrow from v to w indicates $v \not\geq_a w$:



indicates $v \not\geq_a w$

The picture above indicates there is no a-arrow from v to w that is either drawn or implied. So we conclude that $v \not\geq_a w$ only after we have taken into account all drawn and implied a-arrows and determined that no a-arrow from v to w is indicated.

- The picture must specify a relation $\geq_a$ that is *locally connected*: defining for each world w the *connected component*

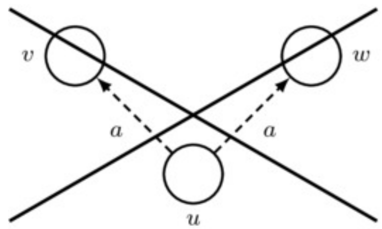
$$\mathsf{cc}_a(w) := \{v \in W \mid w(\geq_a \cup \leq_a)^* v\}$$

of w, we have that $v \in \mathsf{cc}_a(w)$ implies $w \geq_a v$ or $v \geq_a w$.
To explain the meaning of this property, we first explain the definition of the connected component $\mathsf{cc}_a(w)$. This set is based on the union relation $\geq_a \cup \leq_a$, which relates two worlds if and only if they are related according to $\geq_a$ or $\leq_a$; that is, we have $w(\geq_a \cup \leq_a)v$ if and only if $w \geq_a v$ or $w \leq_a v$. We then take the union relation and apply

the operator $(-)^*$, which forms the *reflexive-transitive closure* R^* of a relation R: the relation R^* relates two worlds if and only if they are the same or there exists a sequence of intermediate worlds that are stepwise connected by the underlying relation R. Therefore, we have $w(\geq_a \cup \leq_a)^* v$ if and only if $w = v$ or there exists a sequence $u_1, \dots, u_n$ of worlds such that

$$w(\geq_a \cup \leq_a)u_1(\geq_a \cup \leq_a) \cdots (\geq_a \cup \leq_a)u_n(\geq_a \cup \leq_a)v.$$

So, taken together, we have $v \in \mathsf{cc}_a(w)$ if and only if $v = w$ or there is a sequence $u_1, \dots, u_n$ of worlds connecting v to w stepwise in terms of plausibility (without regard to whether the relative plausibility is increasing, decreasing, or remaining the same). In terms of our pictures of plausibility models, we have $v \in \mathsf{cc}_a(w)$ if and only if, after taking into account all drawn and implied a-arrows and disregarding arrow directionality, v and w are the same or are linked by a sequence of a-arrows (in which each arrow can be followed both forward and backward). The property of local connectedness then tells us that if $v \in \mathsf{cc}_a(w)$, then we must have $w \geq_a v$ or $v \geq_a w$. That is, if there is an undirected a-arrow path from w to v, then there is an a-arrow directly linking w and v in one direction or the other. Therefore, if we think of $\mathsf{cc}_a(w)$ as the set of worlds that the agent a considers to be possible whenever w is the actual world, local connectedness tells us that agent a must always have an opinion as to the relative plausibility between any two worlds that she considers to be possible. It is in this sense that each agent must be "opinionated" as to the relative plausibility of worlds she considers possible. As an example, the property of local connectivity disallows the following picture:



This picture is disallowed because because it violates local connectivity for $\geq_a$: we have $v \in \mathsf{cc}_a(w)$ and yet neither $v \geq_a w$ nor $w \geq_a v$. In detail: we have $v \in \mathsf{cc}_a(w)$ because we have the undirected a-arrow path $w \dashleftarrow_a u \dashrightarrow_a v$ from w to v; and we have neither $y \geq_a z$ nor $z \geq_a y$ because, after adding all implied arrows (in this case only reflexive arrows must be added), we find that we have neither $v \dashrightarrow_a w$ nor $v \dashleftarrow_a w$.

Worlds in the same connected component are said to be *informationally equivalent*.

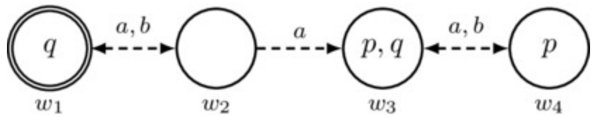
Informational equivalence. Worlds v and w are said to be *informationally equivalent (for agent a)* if and only if $\mathsf{cc}_a(w) = \mathsf{cc}_a(v)$. Notice that we have $\mathsf{cc}_a(w) = \mathsf{cc}_a(v)$ if and only if $v \in \mathsf{cc}_a(w)$ if and only if $w \in \mathsf{cc}_a(v)$.

The idea is that if w is the actual world, then agent a has the information that the actual world must be one of those in her connected component $\mathsf{cc}_a(w)$. Thus the set $\mathsf{cc}_a(w)$ makes up the worlds agent a considers to be possible whenever w is the actual world. And since $w \in \mathsf{cc}_a(w)$, agent a will always consider the actual world to be possible. Local connectivity then guarantees that the agent always has an opinion as to the relative plausibility of any two worlds among those in $\mathsf{cc}_a(w)$ that she considers possible.

One consequence of local connectivity is that informationally equivalent states can be stratified according to Grove's idea (Grove 1988) of concentric spheres: the most plausible worlds overall are positioned on the innermost sphere, the next-most-plausible worlds are positioned on the next-most-larger sphere, and so on, all the way out to the positioning of the least-most-plausible worlds on the largest sphere overall. (The number of worlds in our pictures of plausibility models will always be finite—otherwise we could not draw them according to our above-specified conventions—so it is always possible to organize the worlds in our pictures into concentric spheres in this way.)

Grove spheres (Grove 1988) also suggest a natural method for static belief revision in plausibility models: if the agent is told by a completely trustworthy source that the actual world is among some nonempty subset $S \subseteq \mathsf{cc}_a(w)$ of her informationally equivalent worlds, then she will restrict her attention to the worlds in S. The most plausible worlds in S will be the worlds she then considers to be most plausible overall, the next-most-plausible worlds in S will be the worlds she then considers to be next-most-plausible overall, and so on. That is, she will “reposition” her system of spheres around the set S.

To see how all of this works, let us consider a simple example scenario in which our two agents a and b are discussing the truth of two statements p and q. In the course of the conversation, it becomes common knowledge that neither agent has any information about q and hence neither knows whether q is true, though, as it turns out, q happens to be true. However, it is common knowledge that agent b is an expert about an area of study whose body of work encompasses the question of whether p is true. Further, agent b publicly delivers his expert opinion: p is true. Agent a trusts agent b's expertise and so she (agent a) comes to believe that p is true. But her trust is not absolute: a still maintains the possibility that agent b is wrong or deceitful; hence she is willing to concede that her belief of p is incorrect. Nevertheless, she does trust b for now and comes to believe p. Unfortunately, her trust is misplaced: agent b has knowingly lied; p is actually false. We picture this scenario in Figure 8.



N

Figure 8: The pointed plausibility model (N, w_1).

It is easy to see that the pointed plausibility model (N, w_1) clearly satisfies the property of local connectedness, so this is an allowable picture. To see that this picture reasonably represents the above-describe example scenario, first notice that we have one world for each of the four possible truth assignments to the two letters p and q. At the actual world w_1, the letter p is false and the letter q is true. Agent a considers each of the four worlds to be informationally equivalent (since she does not know with certainty which world is the actual one); however, she considers the p-worlds to be strictly more plausible than the $\neg p$-worlds. This represents her belief that p is true: each of the worlds she considers to be most plausible overall satisfies p. Further, if she is told that p is in fact false, she will restrict her attention to the next-most-plausible $\neg p$-worlds, thereby statically revising her belief. It is in this sense that she trusts b (and so believes p is true) but does not completely rule out the possibility that he is incorrect or deceptive. Since a has no information about q, each of her spheres—the inner p-sphere and the outer $\neg p$-sphere—contains both a world at which q is true and a world at which q is false.

Now let us look at the attitudes of agent b. First, we see that b has two connected components, one consisting of the p-worlds and the other consisting of the $\neg p$-worlds, and these two components are not

informationally equivalent. That is, no p-world is informationally equivalent to a $\neg p$-world in the eyes of agent b. This tells us that b conclusively knows whether p is true. Further, a knows this is so (since each of a's informationally equivalent worlds is one in which b knows whether p is true). Since the actual world is a $\neg p$-world, agent b in fact knows p is false. Finally, we see that b knows that a mistakenly believes that p is true: at each of b's informationally equivalent worlds w_1 and w_2, agent a believes that p is true (since a's most plausible worlds overall, w_3 and w_4, both satisfy p).

We are now ready for the formal definition of plausibility models. This definition summarizes what we have seen so far.

Plausibility model. Given a nonempty set $\mathcal{P}$ of propositional letters and a finite nonempty set $\mathcal{A}$ of agents, a *plausibility model* is a structure

$$M = (W, \geq, V)$$

consisting of

- a nonempty set W of *worlds* identifying the possible states of affairs that might obtain,
- a function $\geq$ that assigns to each agent $a \in \mathcal{A}$ a binary relation $\geq_a$ on W satisfying the property of *Plausibility* that we define shortly, and
- a *propositional valuation* V mapping each propositional letter to the set of worlds at which that letter is true.

We define the following relations, with a negation of a relation indicated by placing a slash through the relational symbol:

- *Converse plausibility*: $w \leq_a v$ means we have $v \geq_a w$.
- *Strict plausibility*: $w >_a v$ means we have $w \geq_a v$ and $v \not\geq_a w$.
- *Strict converse plausibility*: $w <_a v$ means we have $v \geq_a w$ and

$w \not\geq_a v$.

- *Equi-plausibility*: $w \simeq_a v$ means we have $w \geq_a v$ and $v \geq_a w$.

For each world w in W and agent a, we define the *connected component* of w, also called the *a-connected component* if emphasizing a is important, as follows:

$$\mathsf{cc}_a(w) := \{v \in W \mid w(\geq_a \cup \leq_a)^* v\}.$$

If $\mathsf{cc}_a(w) = \mathsf{cc}_a(w)$, then we say that w and v are *informationally equivalent* (or that they are *a-informationally equivalent*). The relation $\geq_a$ must satisfy the property of *Plausibility*, which consists of the following three items:

- $\geq_a$ is reflexive and transitive;
- $\geq_a$ is *locally connected*: $v \in \mathsf{cc}_a(w)$ implies $w \geq_a v$ or $v \geq_a w$; and
- $\geq_a$ is *converse well-founded*: for each nonempty set $S \subseteq W$ of worlds, the set

  $$\min_a(S) := \{w \in S \mid \forall v \in S : v \not<_a w\}$$

  of *a-minimal elements* of S is itself nonempty.

A *pointed plausibility model*, sometimes called a *scenario* or a *situation*, is a pair (M, w) consisting of a plausibility model M and a world w (called the *point*) that designates the state of affairs that we (the formal modelers) currently assume to be actual.

Intuitively, $w \geq_a v$ means that w is no more plausible than v according to agent a. Therefore, it is the "smaller" worlds that are *more* plausible, so that $\min_a(\mathsf{cc}_w(w))$ is the set of worlds that agent a considers to be *most plausible* of all worlds that are informationally equivalent with w.

Local connectivity, as we have seen, ensures that the agent has an opinion as to the relative plausibility of informationally equivalent worlds. Converse well-foundedness guarantees that the agent can always stratify informationally equivalent worlds in such a way that some worlds are the most plausible overall. As a result, we cannot have a situation where agent a has some sequence

$$w_1 >_a w_2 >_a w_3 >_a \cdots$$

of worlds of strictly increasing plausibility, a circumstance in which it would be impossible to find "the most plausible worlds". By forbidding such a circumstance, converse well-foundedness guarantees that the notion of "the most plausible worlds" is always well-defined.

The collection of formulas interpreted on pointed plausibility models generally contains at least the formulas coming from the language $(K\Box)$ defined by the following grammar:

$$\begin{gathered} F ::= p \mid F \wedge F \mid \neg F \mid K_a F \mid \Box_a F \\ p \in \mathcal{P},\ a \in \mathcal{A} \end{gathered} \tag{$K\Box$}$$

The satisfaction relation $\models$ between pointed plausibility models and formulas of $(K\Box)$ is defined as follows.

- $M, w \models p$ holds if and only if $w \in V(p)$.
- $M, w \models F \wedge G$ holds if and only if both $M, w \models F$ and $M, w \models G$.
- $M, w \models \neg F$ holds if and only if $M, w \not\models F$.
- $M, w \models K_a F$ holds if and only if $M, v \models F$ for each $v \in \mathsf{cc}_a(w)$.
- $M, w \models \Box_a F$ holds if and only if $M, v \models F$ for each $v \leq_a w$.

For each $(K\Box)$-formula F and plausibility model $M = (W, \geq, V)$, we define the set

$$[\![F]\!]_M := \{w \in W \mid M, w \models F\}$$

of worlds at which F is true. If M is fixed, we may simply write $[\![F]\!]$ without the subscript M.

K_aF is assigned the reading "agent a has information that F is true". One may consider K_a as a kind of knowledge, though not the kind usually possessed by actual, real-life agents (because it satisfies properties such as closure under logical consequence that are typically not satisfied in practice). Intuitively, possession of information of F is belief in F that persists upon receipt of *any* further information, even information that is *not true*. This kind of knowledge is therefore infallible and indefeasible.

We assign $\Box_aF$ the reading "agent a defeasibly knows F". This is a weak notion of knowledge studied by Lehrer and Paxson (1969) and by Lehrer (1990, 2000) and formalized by Stalnaker (1980, 2006). Intuitively, defeasible knowledge of F is belief in F that persists upon receipt of any further true information: the agent believes F and, if told any further *true* information, she will continue to believe F. Defeasible knowledge is sometimes also called "safe belief".

The dual form of information possession K_aF, written $\hat{K}_aF$, denotes *informational consistency*:

$$\hat{K}_aF \text{ denotes } \neg K_a\neg F.$$

which has the meaning that F is consistent with agent a's information. We use this to define a notion of *conditional belief*:

$$B_a^GF \text{ denotes } \hat{K}_aG \to \hat{K}_a(G \wedge \Box_a(G \to F)).$$

which is assigned the reading "agent a believes F conditional on G". Sometimes B_a^GF is abbreviated by $B_a(F|G)$. Though the meaning of B_a^GF can be derived from the above definitions, the following provides a more intuitive interpretation.

Theorem. For each pointed plausibility model (M, w), we have:

$$M, w \models B_a^GF \quad \text{iff} \quad \min_a([\![G]\!]_M \cap \mathsf{cc}_a(w)) \subseteq [\![F]\!]_M;$$

that is, agent a believes F conditional on G at world w if and only if F is true at the most plausible G-worlds that are consistent with a's information.

This theorem tell us that to see what an agent believes conditional on G, all we need to do is look at the agent's most plausible G-worlds. In this way, conditional belief has the agent "recenter" her system of spheres over the set of all worlds at which G is true. Conditional belief thereby implements static belief revision: to see what agent a believes after statically revising her beliefs by G we simply see what it is she believes conditional on G. Thus B_a^GF says that agent a believes F after statically revising her beliefs by G.

The notion of conditional belief allows us to connect the notions knowledge possession K_a and defeasible knowledge $\Box_a$ with the defeasibility analysis of knowledge, as indicated by the following result.

Theorem. For each pointed plausibility model (M, w), we have each of the following.

- $M, w \models K_aF$ if and only if for each $(K\Box)$-formula G, we have $M, w \models B_a^GF$.
  "Information possession K_a is belief that persists under receipt of *any* information."
- $M, w \models \Box_aF$ if and only if for each $(K\Box)$-formula G satisfying $M, w \models G$, we have $M, w \models B_a^GF$.
  "Defeasible knowledge $\Box_a$ is belief that persists under receipt of *true* information."

Conditional belief gives rise to a notion of *unconditional belief* obtained by taking the trivial condition $\top$ (i.e., the propositional constant for truth) as the condition:

$$B_a F \text{ denotes } B_a^{\top} F.$$

So to see what the agent believes *unconditionally*, we simply conditionalize her beliefs on the trivial condition $\top$, which is true everywhere. It is then easy to see that we have the following.

Theorem. For each pointed plausibility model (M, w), we have:

$$M, w \models B_a F \quad \text{iff} \quad \min_a(\mathsf{cc}_a(w)) \subseteq [\![F]\!]_M;$$

that is, agent a believes F (unconditionally) at world w if and only if F is true at the most plausible worlds that are consistent with a's information.

We conclude this section with the axiomatic theory characterizing those formulas that are valid in all plausibility models. Since we can express conditional belief (and since conditional belief describes static belief revision), what we obtain is a theory of defeasible knowledge, possession of information, conditional belief, unconditional belief, and static belief revision.

The axiomatic theory $\mathsf{K}\Box$.

- The $\mathsf{S5}$ axiom schemes and rules for K_a for each $a \in \mathcal{A}$
- The $\mathsf{S4}$ axiom schemes and rules for $\Box_a$ for each $a \in \mathcal{A}$
- $K_a F \to \Box_a F$
  "If F follows from a's information, then a defeasibly knows F."
- $K_a(\Box_a F \to G) \lor K_a(\Box_a G \to F)$
  "Worlds consistent with the information received are always comparable in terms of plausibility." (That this axiom has this meaning requires some technical details; see Baltag et al. (2014). In particular, the axiom may be viewed as a slight modification of the .3 scheme from basic modal logic; see, e.g., Blackburn et al. 2002).

$\mathsf{K}\Box$ **Soundness and Completeness.** $\mathsf{K}\Box$ is sound and complete with respect to the collection $\mathcal{C}_*$ of pointed plausibility models. That is, for each $(K\Box)$-formula F, we have that $\mathsf{K}\Box \vdash F$ if and only if $\mathcal{C}_* \models F$.

Instead of taking information possession K_a and defeasible knowledge $\Box_a$ as the basic propositional attitudes, one may instead choose conditional belief statements $B_a^G F$. This choice gives the theory CDL of Conditional Doxastic Logic. See Appendix J for details.

We may define a number of additional propositional attitudes beyond conditional belief $B_a^G F$, defeasible knowledge $\Box_a F$, and information possession $K_a F$. We take a brief look at a two of these that have important connections with the Belief Revision literature.

- The *unary revision operator* $*_a F$ has the semantics:

  $$M, w \models *_a F \text{ means } M, w \models F \text{ and } M, v \not\models F \text{ for all } v <_a w$$

  That is, to say that $*_a F$ is true at world w means that F is true at w and that F is false at any world v that agent a ranks strictly more plausible. So to have $*_a F$ true at w says that w will be among the most plausible after the agent undergoes a belief revision by F. The unary revision operator $*_a F$ therefore picks out the worlds that will make up agent a's theory of beliefs after revision by F. As such, we have

  $$M, w \models B_a^F G \quad \text{iff} \quad M, w \models K_a(*_a F \to G),$$

  which says that agent a believes G after revision by F if and only if she knows that G is a consequence of her theory after revision by F.

- For each natural number n, we have a *degree-n belief operator* $B^n_a F$. To define the semantics of these operators, we first define formulas b^n_a for each natural number n according to the following:

$$\begin{aligned} b^0_a &= *_a\top, \\ b^{n+1}_a &= *_a(\neg b^0_a \wedge \neg b^1_a \wedge \cdots \wedge \neg b^n_a). \end{aligned}$$

b^0_a is true at the worlds the agent a ranks most plausible overall, b^1_a is true at the worlds the agent a ranks next-most plausible after the b^0_a-worlds, b^2_a is picks out the next-most plausible after the b^1_a-worlds, and so on. The b^n_a-worlds thereby pick out agent a's "degree-n theory of belief", which is the collection of beliefs the agent will hold after giving up all theories of lesser degree. This setup allows us to realize Spohn's (1988) notion of "degrees of belief":

$$M, w \vDash B^n_a F \text{ means } M, w \vDash K_a(b^n_a \to F) \wedge \bigwedge_{i<n} \neg K_a(b^i_a \to F),$$

which says that agent a believes F to degree n if and only if she knows it follows from her degree-n theory and she does not know it to follow from any of her theories of lesser degree.

4.4 The Logic of Doxastic Actions: Action-priority update

The theories and operators we have seen so far all concern *static* belief change. We now wish to turn to *dynamic* belief change. For this the approach follows the typical pattern in Dynamic Epistemic Logic: we take a given static theory (in this case $\mathsf{K}\Box$) and we add action model-style modalities to create the dynamic theory. When we did this before in the case of basic multi-modal epistemic and doxastic logic, the relational structure of the added action models matched the relational structure of the

models of the theory—Kripke models. The structural match between action models and finite Kripke models is not accidental: the semantics of action model modalities (as explained by the BMS product update) uses the same Kripke model-based notion of agent uncertainty over objects (i.e., the "worlds") to describe agent uncertainty over action model objects (i.e., the "events"). Both uncertainties are represented using the same kind of structure: the binary possibility relation R_a.

For the present theory of conditional belief $B^F_a G$, defeasible knowledge $\Box_a F$, and information possession $K_a F$, we take a similar approach: we define *plausibility action models*, which are action model-type objects whose relational structure matches the relational structure of the models of this theory—plausibility models. Since a finite plausibility model has the form $(W, \geq, V)$, our intuition from the Kripke model case suggests that plausibility action models should have the form $(E, \geq, \mathsf{pre})$, with E a finite nonempty set of events, $\geq$ a function giving a plausibility relation $\geq_a$ for each agent a, and pre a precondition function as before.

Plausibility action model. Given a set of formulas $\mathcal{L}$ and a finite nonempty set $\mathcal{A}$ of agents, a *plausibility action model* is a structure

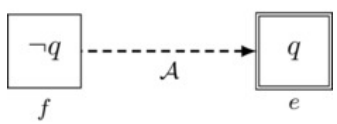
$$A = (E, \geq, \mathsf{pre})$$

consisting of

- a nonempty finite set E of the possible communicative *events* that might occur,
- a function $\geq$ that assigns to each agent $a \in \mathcal{A}$ a binary relation $\geq_a$ on E satisfying the property of Plausibility defined earlier, and
- a function $\mathsf{pre} : E \to \mathcal{L}$ that assigns to each event e in E a formula $\mathsf{pre}(e) \in \mathcal{L}$ called the *precondition* of e. Intuitively, the precondition is announced when the event occurs.

A *pointed plausibility action model*, sometimes also called an *action*, is a pair (A, e) consisting of a plausibility action model A and an event e in A that is called the *point*. In drawing plausibility action models, events are drawn as rectangles, a point (if any) is indicated with a double rectangle, and arrows are drawn using dashes (as for plausibility models). We use many of the same drawing and terminological conventions for plausibility action models that we use for (pointed) plausibility models.

As expected, the main difference between plausibility action models and basic action models is that the agent-specific component (i.e., the function $\geq$ giving the agent-specific relation $\geq_a$). In constructing new plausibility models based on plausibility action models, we may follow a construction similar to the product update. To make this work, our main task is to describe how the plausibility relation $\geq_a$ in the resultant plausibility model $M[A]$ is to be determined in terms of the plausibility relations coming from the given initial plausibility model M and the plausibility action model A. For this it will be helpful to consider an example.



$\mathsf{rPub}_G(q)$

Figure 9: The pointed plausibility action model $(\mathsf{rPub}(q), e)$ for the revisable public announcement of q (also called the "lexicographic upgrade by q" by van Benthem 2007).

Figure 9 depicts $(\mathsf{rPub}(q), e)$, a pointed plausibility action model consisting of two events: the event f in which $\neg q$ is announced and the event e in which q is announced. Event e is the event that actually occurs. For each agent a (coming from the full set of agents $\mathcal{A}$), event e is strictly

more plausible. We adopt the same drawing conventions for plausibility action models that we did for plausibility models: one- and two-way arrows, reflexive and transitive closures, and the requirement of local connectedness. (Well-foundedness follows because the set E of events is finite.) Accordingly, Figure 9 implicitly contains reflexive dashed arrows for each agent at each event.

$(\mathsf{rPub}(q), e)$ has the following intuitive effect: the public announcement of q (i.e., event e) occurs and this is common knowledge; however, the agents still maintain the possibility that the negation $\neg q$ was announced (i.e., event f occurred). In effect, the agents will come to believe q (because the announcement of this was most plausible), but they will nevertheless maintain the less plausible possibility that q is false. This allows the agents to accept the announced formula q but with some caution: they can still revise their beliefs if they later learn that q is false.

The "action-priority update" is the analog of the product update for plausibility models.

Action-priority update (Baltag and Smets 2008b). Let (M, w) be a pointed plausibility model and (A, e) be a pointed plausibility action model. Let $\models$ be a binary satisfaction relation defined between (M, w) and formulas in the language $\mathcal{L}$ of the precondition function $\mathsf{pre}^A : E^A \to \mathcal{L}$ of the plausibility action model A. If $M, w \models \mathsf{pre}^A(e)$, then the plausibility model

$$M[A] = (W[A], \geq[A], V[A])$$

is defined via the *action-priority update* operation $M \mapsto M[A]$ given as follows:

- $W[A] := \{(v, f) \in W \times E \mid M, v \models \mathsf{pre}^A(f)\}$ — pair worlds with events whose preconditions they satisfy;

- $(v_1,f_1) \geq[A]_a (v_2,f_2)$ if and only if we have one of the following:
  - $f_1 >_a f_2$ and $\mathsf{cc}_a(v_1) = \mathsf{cc}_a(v_2)$ — events of strictly differing plausibility are applied to informationally equivalent worlds, or
  - $f_1 \simeq_a f_2$ and $v_1 \geq_a v_2$ — equi-plausible events are applied to informationally equivalent worlds of differing plausibility;
- $V[A]((v,f)) := V^M(p)$ — make the valuation of p at the pair (v,f) just as it was at v.

An action (A,e) operates on an initial situation (M,w) satisfying $M,w \models \mathsf{pre}^A(e)$ via the action-priority update to produce the resultant situation $(M[A],(w,e))$. Note that we may write the plausibility relation $\geq[A]_a$ for agent a after the action-priority update by A simply as $\geq_a$ when the meaning is clear from context.

We now turn to *Action-Priority Update Logic* (a.k.a., *the Logic of Doxastic Actions*). To begin, we define the language (APUL) of Action-Priority Update Logic along with the set PAM_* of pointed plausibility action models having preconditions in the language (APUL) according to the following recursive grammar:

$$\begin{gathered} F ::= p \mid F \wedge F \mid \neg F \mid K_a F \mid \Box_a F \mid [A,e]F \\ p \in \mathcal{P},\ a \in \mathcal{A},\ (A,e) \in \mathsf{PAM}_* \end{gathered} \tag{APUL}$$

The satisfaction relation $\models$ between pointed plausibility models and formulas of (APUL) is the smallest extension of the above-defined satisfaction relation $\models$ for $(K\Box)$ satisfying the following:

- $M,w \models [A,e]G$ holds if and only if $M,w \not\models \mathsf{pre}(e)$ or $M[A],(w,e) \models G$, where the model $M[A]$ is given by the action-priority update.

In addition to the revisable public announcement (Figure 9), there are a number of interesting pointed plausibility action models.

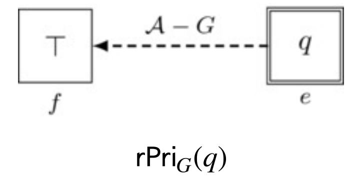


Figure 11: The pointed plausibility action model $(\mathsf{rPri}_G(q), e)$ for the private announcement of q to the group of agents G.

Figure 11 depicts the private announcement of q to a group of agents G. This consists of two events: the event e in which q is announced and the event f in which the propositional constant for truth $\top$ is announced. For agents outside the group G, the most plausible event is the one in which the $\top$ is announced; for agents in the group, the most plausible event is the one in which q is announced. In reality, the announcement of q (i.e., event e) occurs. Since the propositional constant for truth $\top$ is uninformative, the agents outside of G will come to believe that the situation is as it was before. The agents inside G, however, will come to believe q.

The plausibility action model version of the private announcement (Figure 11) is almost identical to the action model version of the private announcement (Figure 3). This is because action models are easily converted into plausibility action models: simply change the arrows to dashed arrows. In this way, we readily obtain plausibility action models from our existing action models. In particular, we can obtain plausibility actions for a public announcement by converting Figure 4, for a semi-private announcement by converting Figure 5, and for a misleading private announcement by converting Figure 6.

Finally, van Benthem (2007) studied two important operations on multi-agent plausibility models that are representable using the action-priority update.

- The *lexicographic upgrade* $[\Uparrow F]G$ (Rott 1989; van Benthem 2007): after changing the plausibility relation so that F-worlds are ranked over $\neg F$-worlds but worlds within each of the F- and $\neg F$-regions are ranked as before, G is true. The lexicographic upgrade by F is just the revisable public announcement of F (Figure 9):

$$M,w \models [\Uparrow F]G \quad \text{iff} \quad M,w \models [\mathsf{rPub}(F),e]G.$$

- The *conservative upgrade* $[\uparrow F]G$ (Boutilier 1993; van Benthem 2007): after changing the plausibility relation so that the best F-worlds are ranked over all other worlds but the rankings are otherwise left unchanged, G is true. We note that in the case where the set $\mathcal{A}$ of agents consists of just the one agent a, we have

$$\begin{aligned} M,w \models [\uparrow F]G \quad &\text{iff} \quad M,w \models [\Uparrow *_a F]G \\ &\text{iff} \quad M,w \models [\mathsf{rPub}(*_a F),e] \end{aligned}$$

which says that the conservative upgrade by F is equal to the lexicographic upgrade by $*_a F$ in the case of a single agent a. This makes sense: $*_a F$ picks out the most plausible F-worlds, and then the lexicographic upgrade $\Uparrow *_a F$ ranks these most plausible F-worlds as most plausible overall and leaves all other rankings as before. In the multi-agent case with n agents, we observe that a plausibility action model with 2^n actions is equivalent to $\Uparrow F$. In particular, let

$$\mathsf{CU}(F) := (E, \geq, \mathsf{pre})$$

be the plausibility action model defined as follows:

◦ $E := \{e_I \mid I \subseteq \mathcal{A}\}$ — there is one event e_I for each (possibly empty) subset $I \subseteq \mathcal{A}$ of agents,

◦ $e_I \geq_a e_J$ if and only if $a \in J$ — event e_J is of equal or grader plausibility according to a iff a is a member of J, and

◦ $\mathsf{pre}(e_I) := (\bigwedge_{i\in I} *_i F) \wedge (\bigwedge_{j\in(\mathcal{A}-I)} \neg *_j F)$ — event e_I picks out the worlds that are best F-worlds for agents in I and not best F-worlds for agents not in I.

Intuitively, this plausibility action model has each agent a split the events into two categories: those that are best F-worlds according to a make up the first category and are ranked highest, and those that are not best F-worlds according to a make up the second category and are ranked strictly less plausible than the first category. In particular, since we have $e_I \leq_a e_J$ if and only if a is in I, it follows that:

◦ if $i \in I \cap J$, then $e_I \simeq_a e_J$;

◦ if $i \in I - J$, then $e_I <_a e_J$;

◦ if $i \in J - I$, then $e_I >_a e_J$;

◦ if $i \in \mathcal{A} - (I \cup J)$, then $e_I \simeq_a e_J$.

So the e_I's having i in I make up the first category and are all ranked most plausible overall by a, and the e_I's not having i in I make up the second category are ranked strictly less plausible by a. The preconditions are arranged so that any two events are pairwise inconsistent: if $I \neq J$, then I and J differ on at least one agent a and therefore they differ on whether their precondition asserts $*_a F$ or its negation $\neg *_a F$. Further, the preconditions of the events exhaust all possibilities: given a world w of a plausibility model, there is a (possibly empty) set of agents I such that the agents in I rank w as a best F-world and the agents not in I do not rank w as a best F-world; as such, world w satisfies the precondition of e_I for the set I in question. Therefore, the events in $\mathsf{CU}(F)$ partition the worlds of a given input plausibility model into a number of pieces, one piece for each subset $I \subseteq \mathcal{A}$. The piece of the given model corresponding to the subset I is picked out by event e_I and consists of those worlds in the given model that satisfy the precondition $\mathsf{pre}(e_I)$; these are the worlds that are best F-worlds according to the agents in I and not best

F-worlds according to the agents not in I. So we see that $\mathsf{CU}(F)$ breaks up the model into various pieces based on which of the agents think the worlds in the given piece are best F-worlds, has the agents rank the pieces so that each agent has her best F-worlds outrank all other worlds, and otherwise leaves the ranking as it was. Accordingly, it is not too hard to see that we have

$$M,w \vDash [\uparrow F]G \quad \text{iff} \quad M,w \vDash \bigvee_{I\subseteq\mathcal{A}} [\mathsf{CU}(F),e_I]G \text{ (general case)},$$

which says that the conservative upgrade by F is equal to the action-priority update brought about by the pointed plausibility action model $(\mathsf{CU}(F),e_I)$ for some subset $I \subseteq \mathcal{A}$. As already stated, a world in the initial model will satisfy the precondition of exactly one event e_J. Therefore, the truth at w of the disjunction $\bigvee_{I\subseteq\mathcal{A}}[\mathsf{CU}(F),e_I]G$ is determined by evaluating the truth at w of the disjunct $[\mathsf{CU}(F),e_J]G$ for the particular J corresponding to w.

We now study the axiomatic theory of Action-Priority Update Logic.

The axiomatic theory APUL.

- Axiom schemes and rules for the theory $\mathsf{K}\Box$
- Reduction axioms:
  1. $[A,e]p \leftrightarrow (\mathsf{pre}(e) \to p)$ for letters $p \in \mathcal{P}$
     "After a non-executable action, every letter holds—a contradiction. After an executable action, letters retain their truth values."
  2. $[A,e](G \wedge H) \leftrightarrow ([A,e]G \wedge [A,e]H)$
     "A conjunction is true after an action iff each conjunct is."
  3. $[A,e]\neg G \leftrightarrow (\mathsf{pre}(e) \to \neg[A,e]G)$
     "G is false after an action iff the action, whenever executable, does not make G true."
  4. $[A,e]K_aG \leftrightarrow (\mathsf{pre}(e) \to \bigwedge_{e\simeq_a f} K_a[A,f]G)$
     "a has information that G after an action iff the action, whenever executable, provides a with information that G will become true despite the uncertainty in her information as to the actual event."
  5. $[A,e]\Box_aG \leftrightarrow (\mathsf{pre}(e) \to (\bigwedge_{e>_a f} K_a[A,f]G)$
     $\wedge (\bigwedge_{e\simeq_a f} \Box_a[A,f]G))$
     "a defeasibly knows G after an action iff the action, whenever executable, provides a with information that G will become true after all more plausible events and, further, gives a defeasible knowledge that G will become true after all equi-plausible events."
- Action Necessitation Rule: from G, infer $[A,e]G$
  "A validity holds after any action."

The first three reduction axioms are identical to the corresponding reduction axioms for EAL. The fourth APUL reduction axiom is almost identical to the fourth EAL reduction axiom. In particular, the fourth EAL reduction axiom, which reads

$$[A,e]K_aG \leftrightarrow (\mathsf{pre}(e) \to \textstyle\bigwedge_{eR_af} K_a[A,f]G),$$

differs only in the conjunction on the right-hand side: the EAL axiom has its conjunction over events related to e via the Kripke model-style relation R_a, whereas the APUL axiom has its conjunction over events related to e via the plausibility model-style relation $\simeq_a$.

The fifth APUL reduction axiom is new. This axiom captures the essence of the action-priority update: for an agent to have defeasible knowledge after an action, she must have information about what happens as a result of more plausible actions and, further, she must have defeasible knowledge about the outcome of equi-plausible actions. The reason for this follows from the definition of the resulting plausibility relation $\geq[A]_a$. As a reminder, this is defined by setting $(v_1,f_1) \geq[A]_a (v_2,f_2)$ if and only if we have one of the following:

- $f_1 >_a f_2$ and $\mathsf{cc}_a(v_1) = \mathsf{cc}_a(v_2)$ — events of strictly differing plausibility are applied to informationally equivalent worlds; or
- $f_1 \simeq_a f_2$ and $v_1 \geq_a v_2$ — equi-plausible events are applied to informationally equivalent worlds of differing plausibility.

Looking to the fifth APUL reduction axiom, the conjunct $\bigwedge_{e>_a f} K_a[A,f]G$ says that G is true whenever an event of plausibility strictly greater than e is applied to a world within a's current connected component. This tells us that G is true at worlds having greater plausibility in light of the first bulleted item above. The other conjunct $\bigwedge_{e\simeq_a f} \Box_a[A,f]G)$ of the fifth APUL reduction axiom says that G is true whenever an event equi-plausible with e is applied to world of equal or greater plausibility within a's current connected component. This tells us that G is true at worlds having greater or equal plausibility in light of the second bulleted item above. Taken together, since these two bulleted items define when it is that a world has equal or greater plausibility in the resultant model $M[A]$, the truth of these two conjuncts at an initial situation (M,w) at which (A,e) is executable implies that G is true at all worlds of equal or greater plausibility than the actual world (w,e) of the resultant model $M[A]$. That is, we have $M[A],(w,e) \models \Box_a G$ and therefore that $M,w \models [A,e]G$. This explains the right-to-left direction of the fifth APUL reduction axiom. The left-to-right direction is explained similarly.

As was the case for EAL, the APUL reduction axioms allow us to "reduce" each formula containing plausibility action models to a provably equivalent formula whose plausibility action model modalities appear before formulas of lesser complexity, allowing us to eliminate plausibility

action model modalities completely via a sequence of provable equivalences. As a consequence, we have the following.

> **APUL Reduction Theorem.** Every F in the language (APUL) is APUL-provably equivalent to a formula F° coming from the plausibility action model-free modal language $(K\Box)$.

Once we have proved APUL is sound, the Reduction Theorem leads us to axiomatic completeness via the known completeness of the underlying modal theory K$\Box$.

> APUL **Soundness and Completeness.** APUL is sound and complete with respect to the collection $\mathcal{C}_*$ of pointed plausibility action models. That is, for each (APUL)-formula F, we have that APUL $\vdash F$ if and only if $\mathcal{C}_* \models F$.

As is the case for EAL, it is possible to combine two consecutive actions into a single action. All that is required is an appropriate notion of plausibility action model composition.

> **Plausibility action model composition.** The *composition* $A \circ B = (E, \geq, \mathsf{pre})$ of plausibility action models $A = (E^A, \geq^A, \mathsf{pre}^A)$ and $B = (E^B, \geq^B, \mathsf{pre}^B)$ is defined as follows:
>
> - $E = E^A \times E^B$ — composed events are pairs (e,f) of constituent events;
> - $(e_1,f_1) \geq_a (e_2,f_2)$ if and only if one of the following obtains:
>   - $e_1 \geq_a e_2$ and $\mathsf{cc}_a(f_1) = \mathsf{cc}_a(f_2)$ — events of differing plausibility are followed by informationally equivalent events, or
>   - $e_1 \simeq_a e_2$ and $f_1 \geq_a f_2$ — equi-plausible events are followed by informationally equivalent events of differing plausibility;

- $\mathsf{pre}((e_1, e_2)) = \mathsf{pre}^A(e_1) \wedge [A, e_1]\mathsf{pre}^B(e_2)$ — a composed event is executable iff the first constituent is executable and, after it occurs, the second constituent is executable as well.

Composition Theorem. Each instance of the following schemes is APUL-derivable.

- Composition Scheme: $[A,e][B,f]G \leftrightarrow [A \circ B,(e,f)]G$
- Associativity Scheme:
  $[A \circ B,(e,f)][C,g]H \leftrightarrow [A,e][B \circ C,(f,g)]H]$

It is also possible to add valuation-changing substitutions (i.e., "factual changes") to plausibility action models. This is done exactly as it is done for action models proper: substitutions are added to plausibility action models, the action-priority update is modified to account for substitutions in the semantics, and the first reduction axiom is changed to account for substitutions in the axiomatics. See Appendix G for details.

4.5 Evidential dynamics and justified belief

One development in DEL is work aimed toward building logics of evidence, belief, and knowledge for use in Formal Epistemology.

- Velázquez-Quesada (2009) and van Benthem and Velázquez-Quesada (2010) study logics of inference and update. These have models containing worlds that explicitly list the formulas of which agents are "aware", like awareness logics (Fagin et al. 1995), except that DEL-style modalities can change these awareness sets, allowing agents to increase the formulas of which they are aware and make inferences with these formulas over time. In this sense, the "awareness sets" may be though of as evidence for the formulas the agents presently know.
- Baltag, Renne, and Smets (2014) study a logic of "conclusive" (or "good") evidence based on a combination of plausibility models with an adaptation of the syntactic bookkeeping mechanisms of Justification Logic (Artemov 2008; Artemov and Fitting 2012). They argue that their work generalizes the awareness logics of Velázquez-Quesada (2009) and van Benthem and Velázquez-Quesada (2010) and better addresses updates with higher-order information.
- A different approach to evidence in Dynamic Epistemic Logic was proposed by van Benthem and Pacuit (2011a,b) and studied further in van Benthem, Fernández-Dunque, and Pacuit (2012, 2014). This approach is much less syntactic than the Justification Logic-style approaches, focusing instead on the semantic notion of modal "neighborhood" (or "minimal") models that have been repurposed with an evidential twist.

We refer the reader to Appendix K for further details.

5. Probabilistic update in Dynamic Epistemic Logic

Dynamic Epistemic Logics that incorporate probability have been studied by a number of authors. Van Benthem (2003), Kooi (2003), Baltag and Smets (2008a), and van Benthem, Gerbrandy, and Kooi (2009b) studied logics of finite probability spaces. Sack (2009) extended the work of Kooi (2003) and van Benthem, Gerbrandy, and Kooi (2009b) to full probability spaces (based on σ-algebras of events). Of these, we mention two in particular:

- Baltag and Smets (2008a) develop logics of finite probability spaces that connect three areas of work: the Popper–Réyni–de Finetti extension of Bayesian probabilistic conditionalization, the theory of Belief Revision, and Dynamic Epistemic Logic. This leads to a definition of an action-model-style probabilistic product update that permits update on events of probability zero (which is required by belief revision).

- Van Benthem, Gerbrandy, and Kooi (2009b) have a different approach with action-model-style probabilistic update that takes into account three sources of probabilistic information: prior probabilities, occurrence probabilities, and observation probabilities.

We refer the reader to Appendix L for further details.

6. Applications of Dynamic Epistemic Logic

6.1 Preference dynamics

DEL-style model-changing operators have been applied by a number of researchers to the study of preferences, preference change, and related notions. We refer the reader to Appendix M for further information, which mentions the work of van Benthem et al. (2009), van Benthem and Liu (2007), Liu (2008), Yamada (2007a,b, 2008), van Eijck (2008), van Eijck and Sietsma (2010), van Benthem, Girard, and Roy (2009c), and Liu (2011).

6.2 Connections with Temporal Logic

Action model-style modalities $[A,e]$ of Dynamic Epistemic Logic have a temporally suggestive reading: "*after* action (A,e), formula F is true". This "before-after" reading suggests, naturally enough, that time passes as actions occur. The semantics of action models supports this suggestion: determining the truth of an action model formula $[A,e]F$ in a model—the model "before" the action—requires us to apply the model-transforming operation induced by the action (A,e) and then see whether F holds in the model that results "after" the action. Channeling Parikh and Ramanujam (2003) some DEL authors further this suggestion by using the temporally charged word "history" to refer to a sequence of pointed Kripke models brought about by the occurrence of a sequence of model-transforming operations. All of this seems to point to the existence of a direct relationship between the occurrence of model-transforming actions and the passage of time: time passes as these actions occur. However, the formal languages introduced so far do not have a built-in means for directly expressing the passage of time, and so, as a consequence, the axiomatic theories developed above are silent on the relationship between the flow of time and the occurrence of model-changing actions. This leaves open the possibility that, within the context of these theories, the passage of time and the occurrence of actions need not necessarily relate as we might otherwise suspect.

For more on this, we refer the interested reader to Appendix N, which mentions a number of studies that bring some method of time-keeping within the scope of the Dynamic Epistemic Logic approach: the work of Sack (2007, 2008, 2010), Yap (2006, 2011), Hoshi (2009), Hoshi and Yap (2009), van Benthem, Gerbrandy, and Pacuit (2007), van Benthem et al. (2009a), Dégremont, Löwe, and Witzel (2011), and Renne, Sack, and Yap (2009, 2015).

6.3 Connections to mainstream Epistemology

A number of works utilize tools and techniques from Dynamic Epistemic Logic for formal reasoning on topics in mainstream Epistemology.

- Baltag and Smets (2008b) use plausibility models (Section 4.3) and The Logic of Doxastic Actions (Section 4.4) to capture a number of notions of knowledge, including Aumann's partition-based notion and Stalnaker's (2006) formalization of Lehrer's (1990, 2000) defeasibility analysis of knowledge.
- Building on the work mentioned in the previous item, Baltag, Renne, and Smets (2014) show that a theory JBG of Justified Belief with

"Good" Evidence can be used to reason about certain examples from mainstream Epistemology. For example, Gettier (1963) constructs a famous counterexample to the claim that "knowledge" may be equated with "justified true belief" (i.e., justified *correct* belief). In this example, an agent—let us call her a—has evidence for a propositional letter f, concludes via logical deduction that $b \vee f$, and therefore has evidence for this disjunction; however, unknown to the agent, f is false but b is true. She therefore has justified true belief but not knowledge that $b \vee f$ is true (since her reason for believing this disjunction is based on her belief in the wrong disjunct). This example is easily reconstructed in JBG, providing one formal account of an agent whose justified belief is correct and yet the agent does not have knowledge (even in a weakly defeasible sense). See Appendix K for additional details.

- Baltag, Renne, and Smets (2012) analyze a Gettier-like example due to Lehrer (1990, 2000) in a variant of JBG that includes dynamic operations for addition of evidence, stepwise logical reasoning, announcement-like addition of evidence with world elimination, and evidence-based plausibility upgrades of worlds.
- Fitch's paradox (Fitch 1963) concerns the seemingly strange result that the existence of unknown truths implies not all truths are knowable. Following the suggestion of van Benthem (2004), Balbiani et al. (2008) equate "knowability" with "being known after some announcement" and show using Arbitrary Public Announcement Logic (see Appendix E) that it is jointly inconsistent to assume that "p is an unknown truth" and that "all truths are knowable". We refer the reader to the discussions in van Benthem (2004), Balbiani et al. (2008), and Brogaard and Salerno (2012) for further details.

7. Conclusion

We have surveyed the literature of Dynamic Epistemic Logic, from its early development in the Public Announcement Logic to the generalized communication operations of action models, work on qualitative and quantitative belief revision, and applications in a variety of areas. Dynamic Epistemic Logic is an active and expanding area, and we have highlighted a number of open problems and directions for further research.

Appendices

A. Kripke models for modal logic
B. Solutions to Cheryl's Birthday, Muddy Children, and Sum and Least Common Multiple
C. Properties of binary relations
D. Normal modal logic
E. Technical details of Public Announcement Logic
F. The axiomatic theories of Relativized Common Knowledge
G. More on action models and the Logic of Epistemic Actions
H. Recursive definition of languages with action models
I. Variants of the action model approach to Dynamic Epistemic Logic
J. Conditional Doxastic Logic
K. Evidential dynamics and justified belief
L. Probabilistic update in Dynamic Epistemic Logic
M. Preference dynamics
N. Temporal aspects of Dynamic Epistemic Logic

Bibliography

Ågotnes, T., P. Balbiani, H. van Ditmarsch, and P. Seban, 2010, "Group announcement logic", *Journal of Applied Logic*, 8(1): 62–81. doi:10.1016/j.jal.2008.12.002

Alchourrón, C.E., P. Gärdenfors, and D. Makinson, 1985, "On the logic of theory change: Partial meet contraction and revision functions", *Journal of Symbolic Logic*, 50(2): 510–530. doi:10.2307/2274239

Andréka, H., I. Németi, and J. van Benthem, 1998, "Modal languages and bounded fragments of predicate logic", *Journal of Philosophical Logic*, 27(3): 217–274. doi:10.1023/A:1004275029985

Areces, C., R. Fervari, and G. Hoffmann, 2012, "Moving arrows and four model checking results", in Ong and Queiroz 2012: 142–153. doi:10.1007/978-3-642-32621-9_11

Arló-Costa, H. and R. Parikh, 2005, "Conditional probability and defeasible inference", *Journal of Philosophical Logic*, 34(1): 97–119. doi:10.1007/s10992-004-5553-6

Artemov, S.N., 2008, "The logic of justification", *The Review of Symbolic Logic*, 1(4): 477–513. doi:10.1017/S1755020308090060

Artemov, S.N. and M. Fitting, 2012, "Justification logic", in E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2012 Edition). URL=<https://plato.stanford.edu/archives/fall2012/entries/logic-justification/>

Aucher, G., 2003, "A combined system for update logic and belief revision", Master's thesis, University of Amsterdam. [Aucher 2003 available online (pdf)]

–––, 2005, "A combined system for update logic and belief revision", in *Intelligent Agents and Multi-Agent Systems: 7th Pacific Rim International Workshop on Multi-Agents, PRIMA 2004, Auckland, New Zealand, August 8–13, 2004, Revised Selected Papers*, Volume 3371 of *Lecture Notes in Computer Science*, pp. 1–17, Berlin/Heidelberg: Springer. doi:10.1007/b107183

–––, 2008, "Consistency preservation and crazy formulas in BMS", in S. Hölldobler, C. Lutz, and H. Wansing (eds.), *Logics in Artificial Intelligence: 11th European Conference, JELIA 2008, Dresden, Germany, September 28–October 1, 2008. Proceedings*, Volume 5293 of *Lecture Notes in Computer Science*, pp. 21–33, Berlin/Heidelberg: Springer. doi:10.1007/978-3-540-87803-2_4

Aucher, G., P. Balbiani, L. Fariñas del Cerro, and A. Herzig, 2009, "Global and local graph modifiers", in C. Areces and S. Demri (eds.), *Proceedings of the 5th Workshop on Methods for Modalities (M4M5 2007)*, Volume 231 of *Electronic Notes in Theoretical Computer Science*, Cachan, France, pp. 293–307. doi:10.1016/j.entcs.2009.02.042

Aucher, G. and F. Schwarzentruber, 2013, "On the complexity of dynamic epistemic logic", in B.C. Schipper (ed.), *Proceedings of the 14th Conference of Theoretical Aspects of Rationality and Knowledge (TARK XIV)*, Chennai, India, pp. 19–28. [Aucher and Schwarzentruber 2013 available online (pdf)]

Balbiani, P., A. Baltag, H. van Ditmarsch, A. Herzig, T. Hoshi, and T. de Lima, 2007, "What can we achieve by arbitrary announcements? A dynamic take on Fitch's knowability", in D. Samet (ed.), *Proceedings of the 11th Conference on Theoretical Aspects of Rationality and Knowledge (TARK XI)*, Brussels, Belgium, pp. 42–51. [Balbiani et al. 2007 available online (pdf)]

–––, 2008, "'Knowable' as 'known after an announcement'", *The Review of Symbolic Logic*, 1(3): 305–334. doi:10.1016/j.entcs.2006.05.034 | [Balbiani et al. 2008 available online (pdf)]

Balbiani, P., H. van Ditmarsch, A. Herzig, and T. De Lima, 2012, "Some truths are best left unsaid", in Bolander et al. 2012: Vol. 9, pp. 36–54. [Balbiani et al. 2012 available online (pdf)]

Ballarin, R., 2010, "Modern origins of modal logic", in E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2014 edition). URL=<https://plato.stanford.edu/archives/win2014/entries/logic-modal-origins/>

Baltag, A. and L.S. Moss, 2004, "Logics for epistemic programs", *Synthese*, 139(2): 165–224. doi:10.1023/B:SYNT.0000024912.56773.5e

Baltag, A., L.S. Moss, and S. Solecki, 1998, "The logic of public announcements, common knowledge, and private suspicions", in I. Gilboa (ed.), *Proceedings of the 7th Conference on Theoretical Aspects of Rationality and Knowledge (TARK VII)*, Evanston, Illinois, USA, pp. 43–56. [Baltag, Moss, and Solecki 1998 available online (pdf)]

–––, 1999, "The logic of public announcements, common knowledge, and private suspicions", Technical Report TR534 (November), Indiana University. [Baltag, Moss, and Solecki 1999 available online (pdf)]

Baltag, A., B. Renne, and S. Smets, 2012, "The logic of justified belief change, soft evidence and defeasible knowledge", in Ong and Queiroz 2012: 168–190. doi:10.1007/978-3-642-32621-9_13

–––, 2014, "The logic of justified belief, explicit knowledge, and conclusive evidence", *Annals of Pure and Applied Logic*, 165(1): 49–81. doi:10.1016/j.apal.2013.07.005

–––, 2015, "Revisable justified belief: Preliminary report", E-print, arXiv.org. arXiv:1503.08141 [cs.LO]. [Baltag, Renne, and Smets 2015 available online (html) | Baltag, Renne, and Smets 2015 available online (pdf)]

Baltag, A. and S. Smets, 2008a, "Probabilistic dynamic belief revision", *Synthese*, 165(2): 179–202. doi:10.1007/s11229-008-9369-8

–––, 2008b, "A qualitative theory of dynamic interactive belief revision", in G. Bonanno, W. van der Hoek, and M. Wooldridge (eds.), *TLG 3: Logic and the Foundations of Game and Decision Theory (LOFT 7)*, Volume 3 of *Texts in logic and games*, pp. 11–58, Amsterdam: Amsterdam University Press. [Baltag and Smets 2008a available online (pdf)]

Blackburn, P., M. de Rijke, and Y. Venema, 2002, *Modal Logic*, Volume 53 of *Cambridge Tracts in Theoretical Computer Science*, Cambridge: Cambridge University Press.

Board, O., 2004, "Dynamic interactive epistemology", *Games and Economic Behavior*, 49(1): 49–80. doi:10.1016/j.geb.2003.10.006

Bolander, T., T. Braüner, S. Ghilardi, and L.S. Moss (eds.), 2012, *Advances in Modal Logic (AiML)*, Copenhagen, Denmark: College Publications.

Boutilier, C., 1993, "Revision sequences and nested conditionals", in *Proceedings of the 13th International Joint Conference on Artificial Intelligence (IJCAI 2011)*, Volume 93, Chambéry, France, pp. 519–531. [Boutilier 1993 available online (pdf)]

–––, 1995, "On the revision of probabilistic belief states", *Notre Dame Journal of Formal Logic*, 36(1): 158–183. doi:10.1305/ndjfl/1040308833

Brogaard, B. and J. Salerno, 2012, "Fitch's paradox of knowability", in E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2013 edition). URL= <https://plato.stanford.edu/archives/win2013/entries/fitch-paradox/>

Bucheli, S., R. Kuznets, J. Sack, and T. Studer, 2010, "Justified belief change", in X. Arrazola and M. Ponte (eds.), *LogKCA-10: Proceedings of the 2nd ILCLI International Workshop on Logic and Philosophy of Knowledge, Communication and Action*, pp. 135–155. The University of the Basque Country Press. [Bucheli et al. available online (pdf)]

Bucheli, S., R. Kuznets, and T. Studer, 2011, "Partial realization in dynamic justification logic", in L.D. Beklemishev and R. de Queiroz (eds.), *Logic, Language, Information and Computation: 18th International Workshop, WoLLIC 2011, Philadelphia, PA, USA, Proceedings*, Volume 6642 of *Lecture Notes in Computer Science*, pp. 35–51, Berlin/Heidelberg: Springer. doi:10.1007/978-3-642-20920-8_9 | [Bucheli et al. 2011 available online (pdf)]

Chang, K., 2015, "A math problem from Singapore goes viral: When is Cheryl's birthday?" *The New York Times* (15 April 2015). [Chang 2015 available online (html)]

Dégrémont, C., B. Löwe, and A. Witzel, 2011, "The synchronicity of dynamic epistemic logic", in K.R. Apt (ed.), *Proceedings of the 13th Conference of Theoretical Aspects of Rationality and Knowledge (TARK XIII)*, Groningen, The Netherlands, pp. 145–152. [Dégrémont et al. 2011 available online (pdf)]

Demey, L., B. Kooi, and J. Sack, 2013, "Logic and probability", in E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2014 edition). URL= <https://plato.stanford.edu/archives/fall2014/entries/logic-probability/>

Fagin, R., J.Y. Halpern, and N. Megiddo, 1990, "A logic for reasoning about probabilities", *Information and Computation*, 87, 78–128. doi:10.1016/0890-5401(90)90060-U

Fagin, R., J.Y. Halpern, Y. Moses, and M. Vardi, 1995, *Reasoning About Knowledge*, Cambridge, MA: MIT Press.

Fervari, R., 2014, *Relation-Changing Modal Logics*, Ph. D. thesis, Facultad de Matemática Astronomía y Física, Universidad Nacional de Córdoba, Córdoba, Argentina.

Fitch, F.B., 1963, "A logical analysis of some value concepts", *The Journal of Symbolic Logic*, 28(2): 135–142.

French, T., W. van der Hoek, P. Iliev, and B. Kooi, 2011, "Succinctness of epistemic languages", in *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI 2011)*, Barcelona, Spain, pp. 881–886. doi:10.5591/978-1-57735-516-8/IJCAI11-153 | [French et al. 2011 available online (pdf)]

–––, 2013, "On the succinctness of some modal logics", *Artificial Intelligence*, 197: 56–85. doi:10.1016/j.artint.2013.02.003

French, T. and H. van Ditmarsch, 2008, "Undecidability for arbitrary public announcement logic", in L. Beklemishev, V. Goranko, and V. Shehtman (eds.), *Advances in Modal Logic (AiML)*, Volume 7, Nancy, France, pp. 23–42. College Publications. [French and van Ditmarsch 2008 available online (pdf)]

Gärdenfors, P., 1988, *Knowledge in Flux: Modeling the Dynamics of Epistemic States*, Cambridge, MA: MIT Press.

Garson, J., 2014, "Modal logic", in E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2014 edition). URL= <https://plato.stanford.edu/archives/sum2014/entries/logic-modal/>

Gerbrandy, J., 1999, *Bisimulations on Planet Kripke*, Ph. D. thesis, University of Amsterdam. [Gerbrandy 1999 available online (pdf)]

Gerbrandy, J. and W. Groeneveld, 1997, "Reasoning about information change", *Journal of Logic, Language and Information*, 6(2): 147–169. doi:10.1023/A:1008222603071

Gettier, E., 1963, "Is justified true belief knowledge?" *Analysis*, 23, 121–123.

Girard, P., 2008, *Modal Logic for Belief and Preference Change*, Ph. D. thesis, University of Amsterdam. [Girard 2008 available online (pdf)]

Girard, P., O. Roy, and M. Marion (eds.), 2011, *Dynamic Formal Epistemology*, Volume 351 of *Synthese Library*, Berlin/Heidelberg: Springer.

Girard, P., J. Seligman, and F. Liu, 2012, "General dynamic dynamic logic", in Bolander et al. 2012: Vol. 8, pp. 239–260. [Girard, Seligman, and Liu 2012 available online (pdf)]

Grove, A., 1988, "Two modellings for theory change", *Journal of Philosophical Logic*, 17(2): 157–170. doi:10.1007/BF00247909

Halpern, J.Y., 2001, "Lexicographic probability, conditional probability, and nonstandard probability", in J. van Benthem (ed.), *Proceedings of the 8th Conference on Theoretical Aspects of Rationality and Knowledge (TARK VIII)*, Certosa di Pontignano, Italy, pp. 17–30. [Halpern 2001 available online (pdf)]

–––, 2003, *Reasoning about Uncertainty*, Cambridge, MA: MIT Press.

Hansson, S.O., 2012, "Logic of belief revision", in E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2014 edition). URL= <https://plato.stanford.edu/archives/win2014/entries/logic-belief-

revision/>

Hendricks, V. and J. Symons, 2006, “Epistemic logic”, in E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2015 edition). URL= <https://plato.stanford.edu/archives/fall2015/entries/logic-epistemic/>

Hintikka, J., 1962, *Knowledge and belief: an introduction to the logic of the two notions*, Cornell: Cornell University Press.

Holliday, W.H., T. Hoshi, and T.F. Icard, III, 2012, “A uniform logic of information dynamics”, in Bolander et al. 2012: Vol. 8, pp. 348–367. [Holliday et al. 2012 available online (pdf)]

Holliday, W.H. and T.F. Icard, III, 2010, “Moorean phenomena in epistemic logic”, in L. Beklemishev, V. Goranko, and V. Shehtman (eds.), *Advances in Modal Logic (AiML)*, Volume 8, Moscow, Russia, pp. 178–199. College Publications. [Holliday and Icard 2010 available online (pdf)]

Hoshi, T., 2008, “Public announcement logics with constrained protocols”, in *Proceedings of the 8th Conference on Logic and the Foundations of Game and Decision Theory (LOFT 8)*, Amsterdam, The Netherlands. [Hoshi 2008 available online (pdf)]

–––, 2009, *Epistemic dynamics and protocol information*, Ph. D. thesis, Stanford University. [Hoshi 2009 available online (pdf)]

–––, 2010, “Merging DEL and ETL”, *Journal of Logic, Language and Information*, 19(4): 413–430. doi:10.1007/s10849-009-9116-7

Hoshi, T. and A. Yap, 2009, “Dynamic epistemic logic with branching temporal structures”, *Synthese*, 169(2): 259–281. doi:10.1007/s11229-009-9552-6

Katsuno, H. and A.O. Mendelzon, 1991, “Propositional knowledge base revision and minimal change”, *Artificial Intelligence*, 52(3): 263–294. doi:10.1016/0004-3702(91)90069-V

Kooi, B., 2003, “Probabilistic dynamic epistemic logic”, *Journal of Logic, Language and Information*, 12(4): 381–408.

doi:10.1023/A:1025050800836

–––, 2007, “Expressivity and completeness for public update logics via reduction axioms”, *Journal of Applied Non-Classical Logics*, 17(2): 231–253. doi:10.3166/jancl.17.231-253

Kooi, B. and B. Renne, 2011a, “Arrow update logic”, *Review of Symbolic Logic*, 4(4): 536–559. doi:10.1017/S1755020311000189

–––, 2011b, “Generalized arrow update logic”, in K.R. Apt (ed.), *Proceedings of the 13th Conference of Theoretical Aspects of Rationality and Knowledge (TARK XIII)*, Groningen, The Netherlands, pp. 205–211. [Kooi and Renne 2011b available online (pdf)]

Kooi, B. and J. van Benthem, 2004, “Reduction axioms for epistemic actions”, in R. Schmidt, I. Pratt-Hartmann, M. Reynolds, and H. Wansing (eds.), *AiML-2004: Advances in Modal Logic (Preliminary Proceedings)*, Manchester, UK, pp. 197–211. [Kooi and van Benthem 2003 available online (pdf)]

Lehrer, K., 1990, *Theory of Knowledge*, Routledge.

–––, 2000, *Theory of Knowledge*, Westview Press.

Lehrer, K. and J. Paxson, T., 1969, “Knowledge: Undefeated justified true belief”, *Journal of Philosophy*, 66, 225–237.

Liu, F., 2008, *Changing for the better: Preference dynamics and agent diversity*, Ph. D. thesis, University of Amsterdam. [Liu 2008 available online (pdf)]

–––, 2010, “Von Wright's ‘The logic of preference’ revisited”, *Synthese*, 175(1): 69–88. doi:10.1007/s11229-009-9530-z

–––, 2011, *Reasoning about preference dynamics*, Volume 354 of *Synthese Library*, Berlin/Heidelberg: Springer. doi:10.1007/978-94-007-1344-4

Lutz, C., 2006, “Complexity and succinctness of public announcement logic”, in *Proceedings of the 5th International Conference on Autonomous Agents and Multiagent Systems (AAMAS-2006)*, Hakodate, Japan, pp. 137–144. Association for Computing Machinery (ACM), New York, New York, USA. doi:10.1145/1160633.1160657

Miller, J.S. and L.S. Moss, 2005, “The undecidability of iterated modal relativization”, *Studia Logica*, 79(3): 373–407. doi:10.1007/s11225-005-3612-9

Ong, L. and R. de Queiroz (eds.), 2012, *Logic, Language, Information and Computation: 19th International Workshop, WoLLIC 2012, Buenos Aires, Argentina, September 3–6, 2012, Proceedings*, Volume 7456 of *Lecture Notes in Computer Science*, Berlin/Heidelberg: Springer-Verlag.

Parikh, R. and R. Ramanujam, 2003, “A knowledge based semantics of messages”, *Journal of Logic, Language and Information*, 12(4): 453–467. doi:10.1023/A:1025007018583

Peppas, P., 2008, “Belief revision”, in F. van Harmelen, V. Lifschitz, and B. Porter (eds.), *Handbook of Knowledge Representation*, Volume 3 of *Foundations of Artificial Intelligence*, Chapter 8, pp. 317–359, Amsterdam, The Netherlands: Elsevier. doi:10.1016/S1574-6526(07)03008-8

Plaza, J., 1989, “Logics of public communications”, in M.L. Emrich, M.S. Pfeifer, M. Hadzikadic, and Z.W. Ras (eds.), *Proceedings of the 4th International Symposium on Methodologies for Intelligent Systems (ISMIS 1989): Poster Session Program*, Charlotte, North Carolina, USA, pp. 201–216. Oak Ridge National Laboratory ORNL/DSRD-24.

–––, 2007, “Logics of public communications”, *Synthese*, 158(2): 165–179. doi:10.1007/s11229-007-9168-7

Popper, K., 2002 [1935], *The Logic of Scientific Discovery*, London, United Kingdom: Routledge Classics. First published in 1935 as *Logik der Forschung*, Vienna: Verlag von Julius Springer.

Renardel de Lavalette, G.R., 2004, “Changing modalities”, *Journal of Logic and Computation*, 14(2): 251–275. doi:10.1093/logcom/14.2.251

Renne, B., 2008, “Public and private communication are different: Results on relative expressivity”, *Synthese*, 165(2): 225–245. doi:10.1007/s11229-008-9395-6

–––, 2009, “Evidence elimination in multi-agent justification logic”, in A. Heifetz (ed.), *Proceedings of the 12th Conference of Theoretical Aspects of Rationality and Knowledge (TARK XII)*, Stanford, California, USA, pp. 227–236. doi:10.1145/1562814.1562845

–––, 2011a, “Public communication in justification logic”, *Journal of Logic and Computation*, 21(6): 1005–1034. doi:10.1093/logcom/exq026

–––, 2011b, “Simple evidence elimination in justification logic”, in Girard et al. 2011: Chapter 7, pp. 127–149. doi:10.1007/978-94-007-0074-1_7

–––, 2012, “Multi-agent justification logic: Communication and evidence elimination”, *Synthese*, 185(S1), 43–82. doi:10.1007/s11229-011-9968-7

Renne, B., J. Sack, and A. Yap, 2009, “Dynamic epistemic temporal logic”, in X. He, J. Horty, and E. Pacuit (eds.), *Logic, Rationality, and Interaction: Second International Workshop, LORI 2009, Chongqing, China, October 8–11, 2009, Proceedings*, Volume 5834 of *Lecture Notes in Computer Science*, pp. 263–277, Berlin/Heidelberg: Springer-Verlag. doi:10.1007/978-3-642-04893-7_21

–––, 2015, “Logics of temporal-epistemic actions”, *Synthese*. doi:10.1007/s11229-015-0773-6. [Renne, Sack, and Yap available online]

Rényi, A., 1955, “On a new axiomatic theory of probability”, *Acta Mathematica Hungarica*, 6(3–4), 285–335. doi:10.1007/BF02024393

–––, 1964, “Sur les espaces simples des probabilités conditionnelles”, in *Annales de l'institut Henri Poincaré (B) Probabilités et Statistiques*,

Volume 1, pp. 3–21. [Rényi 1964 available online]
Roelofsen, F., 2007, “Distributed knowledge”, *Journal of Applied Non-Classical Logics*, 17(2): 255–273. doi:10.3166/jancl.17.255-273
Rott, H., 1989, “Conditionals and theory change: Revisions, expansions, and additions”, *Synthese*, 81(1): 91–113. doi:10.1007/BF00869346
Sack, J., 2007, *Adding Temporal Logic to Dynamic Epistemic Logic*, Ph. D. thesis, Indiana University.
–––, 2008, “Temporal languages for epistemic programs”, *Journal of Logic, Language and Information*, 17(2): 183–216. doi:10.1007/s10849-007-9054-1
–––, 2009, “Extending probabilistic dynamic epistemic logic”, *Synthese*, 169(2): 241–257. doi:10.1007/s11229-009-9555-3
–––, 2010, “Logic for update products and steps into the past”, *Annals of Pure and Applied Logic*, 161(12): 1431–1461. doi:10.1016/j.apal.2010.04.011
Saraf, S. and S. Sourabh, 2012, “Characterizing successful formulas: the multi-agent case”, E-print 1209.0935, arXiv.org. [Saraf and Sourabh 2012 available online (pdf)]
Seligman, J., F. Liu, and P. Girard, 2011, “Logic in the community”, in M. Banerjee and A. Seth (eds.), *Logic and Its Applications: 4th Indian Conference, ICLA 2011, Delhi, India, January 5–11, 2011, Proceedings*, Volume 6521 of *Lecture Notes in Computer Science*, pp. 178–188, Berlin/Heidelberg: Springer. doi:10.1007/978-3-642-18026-2_15
–––, 2013, “Facebook and the epistemic logic of friendship”, in B.C. Schipper (ed.), *Proceedings of the 14th Conference of Theoretical Aspects of Rationality and Knowledge (TARK XIV)*, Chennai, India, pp. 229–238. [Seligman, Liu, and Girard 2013 available online (pdf)]
Sietsma, F. and J. van Eijck, 2012, “Action emulation between canonical models”, in *Proceedings of the 10th Conference on Logic and the Foundations of Game and Decision Theory (LOFT 10)*, Sevilla, Spain. [Sietsma and van Eijck 2012 available online (pdf)]

Sorensen, R., 2011, “Epistemic paradoxes”, in E.N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2014 edition). URL= <https://plato.stanford.edu/archives/spr2014/entries/epistemic-paradoxes/>
Spohn, W., 1988, “Ordinal conditional functions: A dynamic theory of epistemic states”, in W.L. Harper and B. Skyrms (eds.), *Causation in Decision, Belief Change, and Statistics: Proceedings of the Irvine Conference on Probability and Causation*, Volume 42 of *The University of Western Ontario Series in Philosophy of Science*, pp. 105–134, Netherlands: Springer. doi:10.1007/978-94-009-2865-7_6
Stalnaker, R.C., 1980, “A theory of conditionals”, in W.L. Harper, R. Stalnaker, and G. Pearce (eds.), *IFS: Conditionals, Belief, Decision, Chance and Time*, Volume 15 of *The University of Western Ontario Series in Philosophy of Science*, pp. 41–55, Netherlands: Springer. doi:10.1007/978-94-009-9117-0_2
–––, 2006, “On logics of knowledge and belief”, *Philosophical studies*, 128(1): 169–199. doi:10.1007/s11098-005-4062-y
Steiner, D., 2006, “A system for consistency preserving belief change”, in S. Artemov and R. Parikh (eds.), *Proceedings of the Workshop on Rationality and Knowledge, 18th European Summer School in Logic, Language, and Information (ESSLLI)*, Málaga, Spain, pp. 133–144. [Steiner 2006 available online (pdf)]
–––, 2009, *Belief Change Functions for Multi-Agent Systems*, Ph. D. thesis, Universität Bern. [Steiner 2009 available online (pdf)]
Steiner, D. and T. Studer, 2007, “Total public announcements”, in *Logical Foundations of Computer Science: International Symposium, LFCS 2007, New York, NY, USA, June 4–7, 2007, Proceedings*, Volume 4514 of *Lecture Notes in Computer Science*, pp. 498–511, Berlin/Heidelberg: Springer. doi:10.1007/978-3-540-72734-7_35
Steup, M., 2005, “Epistemology”, in E.N. Zalta (ed.), *The Stanford*

Encyclopedia of Philosophy (Spring 2014 edition). URL= <https://plato.stanford.edu/archives/spr2014/entries/epistemology/>

van Benthem, J., 2003, "Conditional probability meets update logic", *Journal of Logic, Language and Information*, 12(4): 409–421. doi:10.1023/A:1025002917675

—–, 2004, "What one may come to know", *Analysis*, 64(2): 95–105. doi:10.1111/j.1467-8284.2004.00467.x

—–, 2005, "An essay on sabotage and obstruction", in D. Hutter and W. Stephan (eds.), *Mechanizing Mathematical Reasoning*, Volume 2605 of *Lecture Notes in Computer Science*, pp. 268–276, Berlin/Heidelberg: Springer. doi:10.1007/978-3-540-32254-2_16

—–, 2007, "Dynamic logic for belief revision", *Journal of Applied Non-Classical Logics*, 17(2): 129–155. doi:10.3166/jancl.17.129-155

—–, 2008a, "For better of for worse: Dynamic logics of preference", Technical Report PP-2008-16, Institute for Logic, Language, Information and Computation (ILLC), University of Amsterdam. [van Benthem 2008a available online (pdf)]

—–, 2008b, "Merging observation and access in dynamic epistemic logic", *Studies in Logic*, 1(1): 1–17.

—–, 2008c, "Merging observation and access in dynamic logic", Technical Report PP-2008-36, Institute for Logic, Language, Information and Computation (ILLC), University of Amsterdam. [van Benthem 2008c available online (pdf)]

van Benthem, J., D. Fernández-Dunque, and E. Pacuit, 2012, "Evidence logic: A new look at neighborhood structures", in Bolander et al. 2012: Vol. 9, pp. 97–118. [van Benthem, Fernández-Dunque, and Pacuit 2012 available online (pdf)]

—–, 2014, "Evidence and plausibility in neighborhood structures", *Annals of Pure and Applied Logic*, 165(1): 106–133. doi:10.1016/j.apal.2013.07.007

van Benthem, J., J. Gerbrandy, T. Hoshi, and E. Pacuit, 2009a, "Merging frameworks for interaction", *Journal of Philosophical Logic*, 38(5): 491–526. doi:10.1007/s10992-008-9099-x

van Benthem, J., J. Gerbrandy, and B. Kooi, 2009b, "Dynamic update with probabilities", *Studia Logica*, 93(1): 67–96. doi:10.1007/s11225-009-9209-y

van Benthem, J., J. Gerbrandy, and E. Pacuit, 2007, "Merging frameworks for interaction: DEL and ETL", in D. Samet (ed.), *Proceedings of the 11th Conference on Theoretical Aspects of Rationality and Knowledge (TARK XI)*, Brussels, Belgium, pp. 43–56. [van Benthem 2007 available online (pdf)]

van Benthem, J., P. Girard, and O. Roy, 2009c, "Everything else being equal: A modal logic for ceteris paribus preferences", *Journal of Philosophical Logic*, 38(1): 83–125. doi:10.1007/s10992-008-9085-3

van Benthem, J. and D. Ikegami, 2008, "Modal fixed-point logic and changing models", in A. Avron, N. Dershowitz, and A. Rabinovich (eds.), *Pillars of Computer Science: Essays Dedicated to Boris (Boaz) Trakhtenbrot on the Occasion of His 85th Birthday*, Volume 4800 of *Lecture Notes in Computer Science*, pp. 146–165, Berlin/Heidelberg: Springer. doi:10.1007/978-3-540-78127-1_9

van Benthem, J. and F. Liu, 2007, "Dynamic logic of preference upgrade", *Journal of Applied Non-Classical Logics*, 17(2): 157–182. doi:10.3166/jancl.17.157-182

van Benthem, J. and E. Pacuit, 2011a, "Dynamic logics of evidence-based beliefs", *Studia Logica*, 99(1–3), 61–92. doi:10.1007/s11225-011-9347-x

—–, 2011b, "Dynamic logics of evidence-based beliefs", Technical Report PP-2011-19, Institute for Logic, Language, Information and Computation (ILLC), University of Amsterdam. [van Benthem and Pacuit 2011b available online (pdf)]

van Benthem, J., J. van Eijck, and B. Kooi, 2006, "Logics of communication and change", *Information and Computation*, 204(11):

1620–1662. doi:10.1016/j.ic.2006.04.006

van Benthem, J. and F.R. Velázquez-Quesada, 2010, "The dynamics of awareness", *Synthese*, 177(1): 5–27. doi:10.1007/s11229-010-9764-9

van Ditmarsch, H., 2005, "Prolegomena to dynamic logic for belief revision", *Synthese*, 147(2): 229–275. doi:10.1007/s11229-005-1349-7

van Ditmarsch, H., J. Eijck, I. Hernández-Antón, F. Sietsma, S. Simon, and F. Soler-Toscano, 2012a, "Modelling cryptographic keys in dynamic epistemic logic with demo", in J.B. Pérez, M.A. Sánchez, P. Mathieu, J.M.C. Rodríguez, E. Adam, A. Ortega, M.N. Moreno, E. Navarro, B. Hirsch, H. Lopes-Cardoso, and V. Julián (eds.), *Highlights on Practical Applications of Agents and Multi-Agent Systems: 10th International Conference on Practical Applications of Agents and Multi-Agent Systems*, Volume 156 of *Advances in Intelligent and Soft Computing*, pp. 155–162, Berlin/Heidelberg: Springer. doi:10.1007/978-3-642-28762-6_19

van Ditmarsch, H. and B. Kooi, 2006, "The secret of my success", *Synthese*, 151(2): 201–232. doi:10.1007/s11229-005-3384-9

van Ditmarsch, H. and J. Ruan, 2007, "Model checking logic puzzles", in *Proceedings of Quatrièmes Journées Francophones Modèls Formels de l'Interaction (MFI'07)*. [van Ditmarsch and Ruan 2007 available online (pdf)]

van Ditmarsch, H., J. Ruan, and R. Verbrugge, 2008, "Sum and product in dynamic epistemic logic", *Journal of Logic and Computation*, 18(4): 563–588. doi:10.1093/logcom/exm081

van Ditmarsch, H., W. van der Hoek, and P. Iliev, 2012b, " Everything is knowable — how to get to know whether a proposition is true", *Theoria*, 78(2): 93–114. doi:10.1111/j.1755-2567.2011.01119.x

van Ditmarsch, H., W. van der Hoek, and B. Kooi, 2005, "Dynamic epistemic logic with assignment", in *Proceedings of the 4th International Conference on Autonomous Agents and Multiagent Systems (AAMAS-2005)*, Utrecht, The Netherlands, pp. 141–148. Association for Computing Machinery (ACM), New York, New York, USA. doi:10.1145/1082473.1082495

–––, 2007, *Dynamic Epistemic Logic*, Volume 337 of *Synthese Library*, Netherlands: Springer. doi:10.1007/978-1-4020-5839-4

van Ditmarsch, H., W. van der Hoek, R. van der Meyden, and J. Ruan, 2006, "Model checking Russian cards", in C. Pecheur and B. Williams (eds.), *Proceedings of the Third Workshop on Model Checking and Artificial Intelligence (MoChArt 2005)*, Volume 149 of *Electronic Notes in Theoretical Computer Science*, pp. 105–123. doi:10.1016/j.entcs.2005.07.029

van Ditmarsch, H., J. van Eijck, and W. Wu, 2010a, "One hundred prisoners and a lightbulb — logic and computation", in F. Lin, U. Sattler, and M. Truszczynski (eds.), *Proceedings of the 12th International Conference on the Principles of Knowledge Representation and Reasoning (KR 2010)*, Toronto, Canada, pp. 90–100. AAAI Press. [van Ditmarsch et al. 2010a available online]

–––, 2010b, "Verifying one hundred prisoners and a lightbulb", *Journal of Applied Non-Classical Logics*, 20(3): 173–191. doi:10.3166/jancl.20.173-191

van Eijck, J., 2005, "Dynamic epistemic modelling", Manuscript version 1.03. [van Eijck 2005 available online (pdf)]

–––, 2008a, "Demo—a demo of epistemic modelling", in J. van Benthem, B. Löwe, and D.M. Gabbay (eds.), *Interactive Logic: Selected Papers from the 7th Augustus de Morgan Workshop, London*, Number 1 in Texts in Logic and Games, pp. 303–362. Amsterdam University Press. doi:10.5117/9789053563564 | [van Eijck 2008a available online (pdf)]

–––, 2008b, "Yet more modal logics of preference change and belief revision", in K.R. Apt and R. van Rooij (eds.), *New Perspectives on Games and Interaction*, Volume 4 of *Texts in Logic and Games*, pp.

81–104. Amsterdam University Press. [van Eijck 2008b available online (pdf)]

van Eijck, J. and S. Orzan, 2005, “Modelling the epistemics of communication with functional programming”, in M. van Eekelen (ed.), *Proceedings of the 6th Symposium on Trends in Functional Programming (TFP 2005)*, pp. 44–59. [van Eijck and Orzan 2005 available online (pdf)]

van Eijck, J., J. Ruan, and T. Sadzik, 2012, “Action emulation”, *Synthese*, 185(1): 131–151. doi:10.1007/s11229-012-0083-1

van Eijck, J. and F. Sietsma, 2010, “Multi-agent belief revision with linked preferences”, in G. Bonanno, B. Löwe, and W. Hoek (eds.), *Logic and the Foundations of Game and Decision Theory — LOFT 8: 8th International Conference, Amsterdam, The Netherlands, July 3–5, 2008, Revised Selected Papers*, Volume 6006 of *Lecture Notes in Computer Science*, pp. 174–189, Berlin/Heidelberg: Springer. doi:10.1007/978-3-642-15164-4_9

van Eijck, J. and Y. Wang, 2008, “Propositional dynamic logic as a logic of belief revision”, in W. Hodges and R. de Queiroz (eds.), *Logic, Language, Information and Computation: 15th International Workshop, WoLLIC 2008 Edinburgh, UK, July 1–4, 2008 Proceedings*, Volume 5110 of *Lecture Notes in Computer Science*, pp. 136–148, Berlin/Heidelberg: Springer. doi:10.1007/978-3-540-69937-8_13

van Fraassen, B.C., 1976, “Representational of conditional probabilities”, *Journal of Philosophical Logic*, 5(3): 417–430.

–––, 1995, “Fine-grained opinion, probability, and the logic of full belief”, *Journal of Philosophical Logic*, 24(4): 349–377. doi:10.1007/BF01048352

Velázquez-Quesada, F.R., 2009, “Inference and update”, *Synthese*, 169(2): 283–300. doi:10.1007/s11229-009-9556-2

Visser, A., 1994, “Actions under Presuppositions”, in J. van Eijck and A. Visser (eds.), *Logic and Information Flow*, pp. 196–233, Cambridge, MA: MIT Press.

von Wright, G.H., 1963, *The Logic of Preference*, Edinburgh University Press.

Wang, Y. and Q. Cao, 2013, “On axiomatizations of public announcement logic”, *Synthese*, 190(1): 103–134. doi:10.1007/s11229-012-0233-5

Wáng, Y.N. and T. Ågotnes, 2011, “Public announcement logic with distributed knowledge”, in H. Ditmarsch, J. Lang, and S. Ju (eds.), *Logic, Rationality, and Interaction: Third International Workshop, LORI 2011, Guangzhou, China, October 10–13, 2011, Proceedings*, Volume 6953 of *Lecture Notes in Computer Science*, pp. 328–341, Berlin/Heidelberg: Springer. doi:10.1007/978-3-642-24130-7_24

–––, 2013, “Public announcement logic with distributed knowledge: expressivity, completeness and complexity”, *Synthese*, 190(1): 135–162. doi:10.1007/s11229-012-0243-3

Yamada, T., 2007a, “Acts of commanding and changing obligations”, in K. Inoue, K. Satoh, and F. Toni (eds.), *Computational Logic in Multi-Agent Systems: 7th International Workshop, CLIMA VII, Hakodate, Japan, May 8–9, 2006, Revised Selected and Invited Papers*, Volume 4371 of *Lecture Notes in Computer Science*, pp. 1–19, Berlin/Heidelberg: Springer. doi:10.1007/978-3-540-69619-3_1

–––, 2007b, “Logical dynamics of some speech acts that affect obligations and preferences”, in J. van Benthem, S. Ju, and F. Veltman (eds.), *A Meeting of the Minds: Proceedings of the Workshop on Logic, Rationality and Interaction, Beijing, 2007 (LORI-I)*, Volume 8 of *Texts in Computer Science*, pp. 275–290. College Publications.

–––, 2008, “Logical dynamics of some speech acts that affect obligations and preferences”, *Synthese*, 165(2): 295–315. doi:10.1007/s11229-008-9368-9

Yap, A., 2006, “Product update and looking backward”, Technical Report PP-2006-39, Institute for Logic, Language, Information and

Computation (ILLC), University of Amsterdam. [Yap 2006 available online (pdf)]

——, 2011, "Dynamic epistemic logic and temporal modality", in Girard et al. 2011: Chapter 3, pp. 33–50. doi:10.1007/978-94-007-0074-1_3

Academic Tools

- How to cite this entry.
- Preview the PDF version of this entry at the Friends of the SEP Society.
- Look up this entry topic at the Indiana Philosophy Ontology Project (InPhO).
- Enhanced bibliography for this entry at PhilPapers, with links to its database.

Other Internet Resources

[Please contact the author with suggestions.]

Related Entries

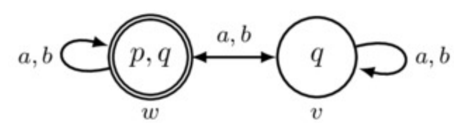
belief, formal representations of | common knowledge | epistemic paradoxes | epistemology | Fitch's paradox of knowability | logic, history of: modal logic | logic: and probability | logic: classical | logic: conditionals | logic: epistemic | logic: justification | logic: modal | logic: of belief revision | logic: temporal

Acknowledgments

Bryan Renne was supported by Innovational Research Incentives Scheme Veni grant 275-20-030 from the Netherlands Organisation for Scientific Research (NWO). The grant was hosted by the Institute for Logic, Language, Information and Computation (ILLC) at the University of Amsterdam.

Appendix A: Kripke models for modal logic

Kripke's models for modal logic (or variants thereof) are the basis for many modern approaches to reasoning about knowledge and belief (Fagin et al. 1995). Intuitively, these models characterize the knowledge (or beliefs) of idealized agents in terms of *considered possibilities*: to say that an agent knows (or believes) some statement F means that F holds in all of the states of affairs that the agent considers as possible candidates for the actual state of affairs. Hence eliminating from consideration one or more alternative states of affairs generally increases one's certainty, thereby trending toward increased knowledge (or belief). In contrast, entertaining additional (perhaps new) states of affairs generally decreases one's certainty, thereby trending toward decreased knowledge (or belief). Higher-order knowledge or belief (i.e., knowledge [or beliefs] about knowledge [or beliefs]) is represented by *iterated considered possibilities*; for example, if F is true of each possibility that agent b entertains when in any one of the states of affairs that agent a entertains, then we say that a knows (or believes) b knows (or believes) F. To make these concepts more concrete, let us consider a simple example.



M

Figure A1. A Kripke model.

Figure A1 pictures a Kripke model M. This model consists of two worlds: an "actual" world w (indicated by the double circle) in which the propositional letters p and q are true and an alternative world v in which p is false and q is true. The actual world indicates what is in fact true of the situation being modeled; in this case, p and q are true. Alternative worlds serve as candidates for the actual world that the agents may consider. As for which worlds are considered, an agent-labeled arrow pointing from a world x to a world y indicates that the agent in question will consider world y as a possible candidate for the actual state of affairs whenever we assume that world x is actual. Therefore, as we can see from the picture, each of agents a and b entertains both worlds w and v as possible states of affairs: a-arrows point from w to w, from w to v, from v to w, and from v to v; and b-arrows also point between the same pairs of worlds. Adopting an epistemic (i.e., knowledge-based) reading for the purposes of this example, an agent in the state of affairs given by world x is by definition said to *know a statement* F if and only if F is true in all worlds the agent considers possible with respect to a (i.e., those worlds that are the target of an arrow that has source x and that is labeled by that agent's name). Looking at the picture, we see that neither agent knows that p is true. Indeed, each agent entertains a possibility—world v—in which p is false and therefore does not know that p is true. Likewise, neither agent knows that p is false because each entertains a world—world w—in which p is true. Defining the phrase "knowing whether F is true" to mean "knowing that F is true or knowing that F is false", what we have seen is that neither agent knows *whether* p is true. Taking our analysis further, let us now show that each agent knows that neither agent knows whether p is true, a statement of *higher-order knowledge* (i.e., knowledge about knowledge). Indeed, in each of the possibilities agent a considers (i.e., in both w and v), neither agent b nor agent a herself knows whether p is true, as we have already seen. Therefore, since it is true in each of agent a's possibilities that "neither agent knows whether p is true", it follows that agent a knows that "neither agent knows whether p is true". Similar reasoning leads us to conclude that agent b knows that "neither agent knows whether p is true".

Hence we conclude that each agent knows that "neither agent knows whether p is true", as was claimed. This kind of analysis can be taken even further to show that it is in fact *common knowledge* among agents a and b that neither knows whether p is true; that is, neither agent knows whether p is true, each agent knows neither agent knows whether p is true, each agent knows each agent knows neither agent knows whether p is true, and so on, taking any finite number of iterations of the phrase "each agent knows" followed by the phrase "neither agent knows whether p is true". One can also undertake a further analysis to show that it is common knowledge that q is true.

Formally, the notion of Kripke model is defined as follows.

> **Kripke model.** Given a nonempty set $\mathcal{P}$ of propositional letters and a finite nonempty set $\mathcal{A}$ of agents, a *Kripke model* is a structure
>
> $$M = (W, R, V)$$
>
> consisting of
>
> - a nonempty set W of *worlds* identifying the possible states of affairs that might obtain,
> - a function $R : \mathcal{A} \to \wp(W \times W)$ that assigns to each agent a a binary *possibility relation* $R_a \subseteq W \times W$ with $wR_a v$ indicating that agent a will entertain world v as a candidate for the actual world whenever we assume that w is in fact actual), and
> - a *propositional valuation* $V : \mathcal{P} \to \wp(W)$ mapping each propositional letter $p \in \mathcal{P}$ to the set $V(p) \subseteq W$ of worlds at which that letter is true.
>
> Notation: if M is a Kripke model, then adding a superscript M to a symbol in $\{W, R, V\}$ is used to denote a component of the triple that makes up M in such a way that $(W^M, R^M, V^M) = M$. An *a-arrow*

(sometimes also called an R_a-*arrow*) is a pair of worlds (w, v) satisfying $wR_a v$, which says that the a-arrow points from the source world w to the target world v. A *pointed Kripke model*, sometimes called a *scenario* or a *situation*, is a pair (M, w) consisting of a Kripke model M and a world $w \in W^M$ (called the *point*) that designates the state of affairs that we (the formal modelers) currently assume to be actual. In drawing a Kripke model, worlds are indicated as circles labeled by their names, a letter p is true at a world w (i.e., $w \in V(p)$) if and only if p is listed inside the circle for world w, and arrows between worlds are drawn explicitly, often with a single arrow labeled by a set or a list of agents denoting a family of arrows, one for each agent in the set or list. If the model is pointed, the point is indicated by a double circle. See Figure A1 above for an example drawing.

For a given possibility relation R_a, some combination of relational properties such as reflexivity, transitivity, Euclideanness, symmetry, and seriality (see Appendix C for definitions), among others, may be assumed based on the particular cognitive notion that this R_a is supposed to represent and on one's philosophical position as to the properties satisfied by the notion in question. The traditional "logic of knowledge" assumes reflexivity, transitivity, and Euclideanness; the traditional "logic of belief" assumes seriality, transitivity, and Euclideanness (Fagin et al. 1995).

The collection of formulas interpreted on pointed Kripke models generally contains at least the formulas coming from a modal language (ML) defined by the following grammar:

$$\begin{gathered} F ::= p \mid F \wedge F \mid \neg F \mid [a]F \\ p \in \mathcal{P},\ a \in \mathcal{A} \end{gathered} \tag{ML}$$

In words, the above "Backus–Naur Form" for defining symbolic notation defines the formulas of the language (ML) (a.k.a., the "(ML)-formulas") as follows. First, each of the letters p coming from a given nonempty set $\mathcal{P}$

of propositional letters is an (ML)-formula. Second, given (ML)-formulas F_1 and F_2, each of the following is also an (ML)-formula: the conjunction $F_1 \wedge F_2$, the negation $\neg F_1$, and the modal expression $[a]F_1$ obtained using an a coming from a given finite (and nonempty) set $\mathcal{A}$ of agent names. (The symbol "$[a]$" is called a *modal operator*, a *modality*, or a *modal*.) Third, something is an (ML)-formula if and only if it can be formed from a propositional letter by way of conjunction, negation, or addition of a modal in the manner just specified. And this defines the collection of (ML)-formulas. When convenient, we will identify the expression "(ML)" with the set of (ML)-formulas. Other languages we define later will also be defined by Backus–Naur notation using similar notational conventions.

We make use of the usual notational abbreviations (from classical propositional logic) when writing formulas containing Boolean connectives other than conjunction ($\wedge$) or negation ($\neg$); this includes, among others, appropriate abbreviations for material implication ($\to$), disjunction ($\vee$), the always-true propositional constant $\top$ for truth, and the always-false propositional constant $\bot$ for falsehood. The formula $[a]F$ is assigned either the epistemic reading "agent a knows F" or the doxastic reading "agent a believes F", with the particular choice of reading depending on the situation to be modeled. Formulas are therefore used to express Boolean combinations of "ontic" (i.e., non-epistemic and non-doxastic) statements along with statements of agents' knowledge (or beliefs). As an example, the formula $p \wedge \neg[a]p$, which says that "p is true and agent a does not know that p is true", expresses an ontic fact p along with the negation of the epistemic statement $[a]p$. (Here an epistemic reading of $\neg[a]p$ was taken; a doxastic one could have been used just as well.) For a set B of agents, we make the following abbreviation:

$$[B]F \text{ denotes } \bigwedge_{a \in B} [a]F.$$

which is the conjunction of all formulas $[a]F$ with a ranging over the set B. The formula $[B]F$ therefore says, "everyone in group B knows (or believes) F". If common knowledge is to be included in the language for the purposes of a particular scenario, then the grammar (ML+C) is used:

$$F ::= p \mid F \wedge F \mid \neg F \mid [a]F \mid [B*]F \qquad \text{(ML+C)}$$
$$p \in \mathcal{P},\ a \in \mathcal{A},\ B \subseteq \mathcal{A}$$

The formula $[B*]F$ is read epistemically as "F is common knowledge among group B" and doxastically as "F is common belief among group B". The expression $[C]F$ is used as an abbreviation for $[\mathcal{A}]F$ ("it is common knowledge that F is true" or "it is common belief that F is true"). Sometimes authors choose to restrict grammar (ML+C) by requiring that the common knowledge group is always the full group of agents (i.e., $B = \mathcal{A}$), but we will not do this here.

Each of the modal operators written between square brackets has a "dual" form defined by way of abbreviation and written using angled brackets:

- The dual $\langle a\rangle$ of $[a]$ is an abbreviation defined by setting $\langle a\rangle F := \neg[a]\neg F$.
- Given a set B of agents, the dual $\langle B\rangle$ of $[B]$ is an abbreviation defined by setting $\langle B\rangle F := \neg[B]\neg F$.
- Given a set B of agents, the dual $\langle B*\rangle$ of $[B]$ is an abbreviation defined by setting $\langle B*\rangle F := \neg[B*]\neg F$.

The dual form of an operator expresses consistency. For example, $\langle a\rangle F$ says, "F is consistent with agent a's beliefs". A knowledge reading is also possible.

Some languages extend (ML) by addition of the *universal modality* $\forall$. The formula $[\forall]F$ says that F is true at every world in the Kripke model. The dual of the universal modality is the *existential modality*, written for convenience as $[\exists]$ and defined as the abbreviation $[\exists]F := \neg[\forall]\neg F$. The formula $[\exists]F$ says that F is true at some world in the model.

The binary truth relation $\models$ between pointed models (written without delimiting parentheses) and formulas is defined inductively as follows:

- $M, w \models p$ holds if and only if $w \in V(p)$.
- $M, w \models F \wedge G$ holds if and only if $M, w \models F$ and $M, w \models G$.
- $M, w \models \neg F$ holds if and only if $M \not\models F$.
- $M, w \models [a]F$ holds if and only if $M, v \models F$ for each v satisfying $wR_a v$.
- $M, w \models [B]F$ holds if and only if $M, v \models F$ for each v satisfying $wR_B v$.
- $M, w \models [B*]F$ holds if and only if $M, v \models F$ for each v satisfying $w(R_B)^* v$.
- $M, w \models [\forall]F$ holds if and only if $M, v \models F$ for each $v \in W$.

The relation $R_B := \bigcup_{a\in A} R_a$ is the union of all relations R_a with a in B, so that $xR_B y$ holds if and only if $xR_a y$ holds for some a in B. The relation $(R_B)^*$ is the *reflexive-transitive closure* of R_B, so that $x(R_B)^* y$ holds if and only if $x = y$ or there is a finite sequence

$$xR_B z_1 R_B z_2 R_B \cdots R_B z_n R_B y$$

of R_B-arrows connecting x to y. Said roughly, R_B says what is considered possible by any one of the members of group B, and $(R_B)^*$ says what is jointly considered possible by members of group B when the members' possibilities are chained together (i.e., one thinks something is possible, which leads another to think something else possible, leading another to consider yet another possibility, and so on, somewhat like a group of paranoid worriers sequentially raising more and more far-fetched possible ways something might go wrong). We write $M, w \models G$ to mean that formula G is true at world w of the pointed model (M, w); the negation is written $M, w \not\models G$. As an exercise, one can go back to Figure A1 and check that

- $M, w \models p \wedge [a]p$ — "p is true at w but agent a does not know it"
- $M, w \models [b]\neg[a]p$ — "agent b knows that agent a does not know p"; and
- $M, w \models [C]\neg([a]p \vee [a]\neg p)$ — "it is common knowledge that agent a does not know whether p is true".

If $\mathcal{C}_*$ is a collection of pointed Kripke models, then $\mathcal{C}_* \models G$ means that G is true at each pointed model in collection $\mathcal{C}_*$. Finally, $\models G$ means that G is true at all pointed Kripke models. For example, one can readily verify that

$$\models [a](F \to G) \to ([a]F \to [a]G), \qquad \text{(K)}$$

which says that agent a's knowledge is closed under the classical rule of *Modus Ponens*: if agent a knows an implication and she knows its antecedent, then she also knows its consequent. The same of course holds for the doxastic reading of the modal operator $[a]$.

For more on Kripke models, we refer the reader to the following entries in the *Stanford Encyclopedia of Philosophy*: Modal Logic, Modern Origins of Modal Logic, and Epistemic Logic.

Appendix B: Solutions to Cheryl's Birthday, Muddy Children, and Sum and Least Common Multiple

---

1. Cheryl's Birthday

1.1 Problem Statement

Cheryl's Birthday (version of Chang (2015, 15 April)). Albert and Bernard just met Cheryl. "When's your birthday?" Albert asked Cheryl. Cheryl thought a second and said, "I'm not going to tell you, but I'll give you some clues". She wrote down a list of 10 dates:

- May 15, May 16, May 19
- June 17, June 18
- July 14, July 16
- August 14, August 15, August 17

"My birthday is one of these", she said. Then Cheryl whispered in Albert's ear the month—and only the month—of her birthday. To Bernard, she whispered the day, and only the day. "Can you figure it out now?" she asked Albert.

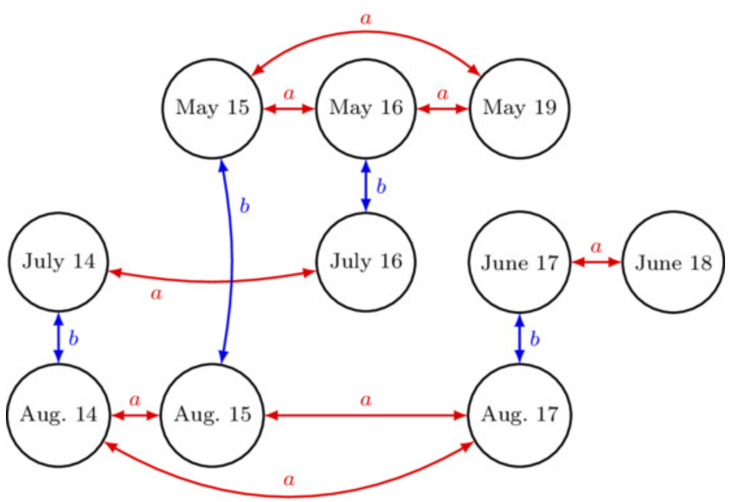
Albert: I don't know when your birthday is, but I know Bernard doesn't know either.
Bernard: I didn't know originally, but now I do.

Albert: Well, now I know too!

When is Cheryl's birthday?

1.2 Solution

We represent Albert's and Bernard's knowledge just after Cheryl asks "Can you figure it out now?" but before Albert and Bernard have their conversation using the model M_{cb} pictured in Figure B1.



M_{cb}

Figure B1. The Kripke model M_{cb}: the situation just after Cheryl asks "Can you figure it out now?" but before Albert and Bernard have their conversation. Red a-arrows represent Albert's uncertainty, and blue b-arrows represent Bernard's uncertainty. For simplicity, the picture omits reflexive arrows for each agent at each date (i.e., to complete the picture, we must add at each world both an a-arrow and a b-arrow pointing in a loop from the date in question back to itself).

In this model, we have one possible world for each date. A red a-arrow pointing from date D to date D' tells us the following: "If Cheryl's birthday is in fact D, then Albert considers it possible that her birthday might be D'". For example, there is a red a-arrow pointing from June 17 to June 18: assuming Cheryl's birthday is in fact June 17, then Albert will consider it possible that her birthday is June 18 because he was told the month—here assumed to be June—so if the date is actually June 17, then he considers June 18 a possibility. Note that there should also be a red a-arrow from June 17 to itself; indeed, if the actual date is June 17, then Albert should consider both June 18 and June 17 to be possible. However, to simplify the picture, we have omitted such "reflexive" arrows (i.e., arrows pointing in a loop from one date back to that very same date). Properly speaking, these arrows should be added in order to complete the picture: every date should have a reflexive red a-arrow.

Just as red a-arrows indicate the various possibilities Albert considers, blue b-arrows tell us the possibilities Bernard considers: a blue b-arrow from date D to date D' tells us that "If Cheryl's birthday is in fact D, then Bernard considers it possible that her birthday might be D'". Reflexive blue b-arrows have also been omitted from the picture for simplicity.

It is not too difficult to verify that the arrow structure in Figure B1 correctly describes the situation with regard to Albert's and Bernard's knowledge just after Cheryl asks "Can you figure it out now?" but before Albert and Bernard have their conversation. Indeed, according to the usual Kripke-style notion of knowledge (i.e., "an agent knows F if and only if F is true at all worlds he considers possible"), Figure B1 shows that Albert knows the month: there are red a-arrows pointing between all dates having the same month. Therefore, no matter which of these dates is Cheryl's birthday, all possibilities Albert will entertain will have the same month and therefore he can be said to know the month of her birthday. Figure B1 also shows that Bertrand knows the numerical day: there are blue b-arrows pointing between all dates having the same numerical day. So no matter which of these dates is Cheryl's birthday, all possibilities Bertrand will entertain will have the same day and therefore he can be said to know the day of her birthday.

Now let us consider the effect of Albert's first announcement, which we label as A_1:

(A_1) Albert: I don't know when [Cheryl's] birthday is, but I know Bernard doesn't know either.

The effect of this statement is a public announcement: we eliminate all dates at which this statement is false, thereby retaining only those dates at which it is true. That is, we eliminate all those possibilities at which

1. Albert knows the birthday, or
2. Albert knows Bernard knows the birthday.

For one of our agents Albert or Bernard to "know" the birthday, he must consider only one of the dates to be possible. That is, supposing date D is the birthday, then an agent will know that "D is Cheryl's birthday" if and only if there is no arrow for that agent pointing from D to some different date D'. So, looking at the dates in Figure B1, we can see that there is no date at which Albert knows the birthday: at each date D, there is a red a-arrow pointing to some different date D'. Therefore, the first of Albert's first public announcement ("I don't know when [Cheryl's] birthday is") does not call for us to eliminate any dates.

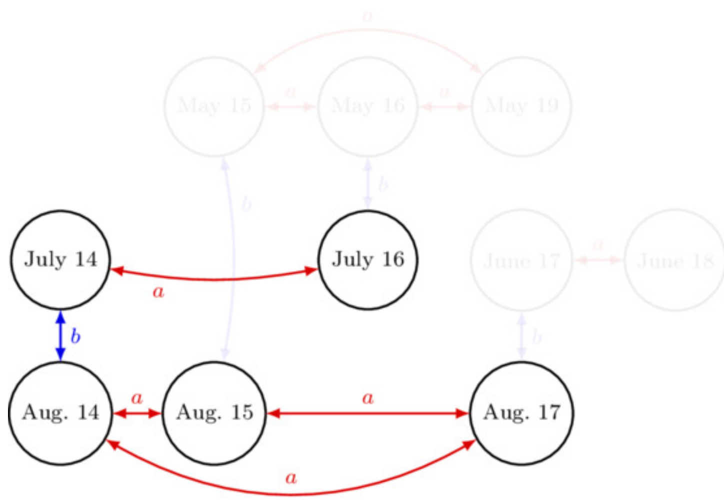
Now we return to the second part of Albert's first public announcement: "I know Bernard doesn't know [Cheryl's birthday] either". This calls for us to eliminate from Figure B1 any date at which this statement is false: Albert does *not* know that Bernard does not know the birthday. To understand which dates we should eliminate, let us look at this in stages.

First, if date D is the birthday and Albert does not know some statement F, then there must be a red a-arrow from D to some date D' (which might be the same as D) and F must be false at D'. Second, the statement F we are interested in is "Bernard does not know the birthday". So for this statement F to be *false*, it must be that Bernard *does* know the birthday: there must be no blue b-arrow pointing to a different date. So since we must eliminate all dates D at which Albert does not know that Bernard knows the birthday, we must eliminate D if there is a red a-arrow pointing to a not necessarily different date D' at which there are no blue b-arrows pointing to a different date. Note: it is important to take into account the reflexive red a-arrows that have been omitted from the picture; this is why D' can be the same as D.

Looking to Figure B1, at which dates can we take a red a-arrow (including the possibility of a non-drawn reflexive a-arrow, which loops us right back to the same date) to a location at which there is no blue b-arrow pointing to a different date? These are:

- May 15: from this date, we can take an a-arrow to May 19, which has no b-arrows to a different date;
- May 16: same as for May 15;
- May 19: from this date, we can take a reflexive a-arrow (not drawn) right back to May 19, which has no b-arrows to a different date;
- June 17: from this date, we can take an a-arrow to June 18, which has no b-arrows to a different date; and
- June 18: from this date, we can take a reflexive a-arrow (not drawn) right back to June 18, which has no b-arrows to a different date.

These are the dates we should eliminate. All other dates should remain: at these other dates, every a-arrow (including the reflexive ones) leads us to a location at which there is indeed a b-arrow going to a different date. So after elimination, we obtain the result of Albert's first public announcement A_1: the model $M_{cb}[A_1]$ pictured in Figure B2.



$M_{cb}[A_1]$

FIGURE B2. The model $M_{cb}[A_1]$ obtained from M_{cb} (Figure B1) after Albert's first public announcement A_1. Eliminated worlds and arrows from the model M_{cb} are faintly drawn in this picture for clarity; these are not actually present in the model. Also, as with Figure B1, reflexive arrows for each agent at each world have been omitted for simplicity.

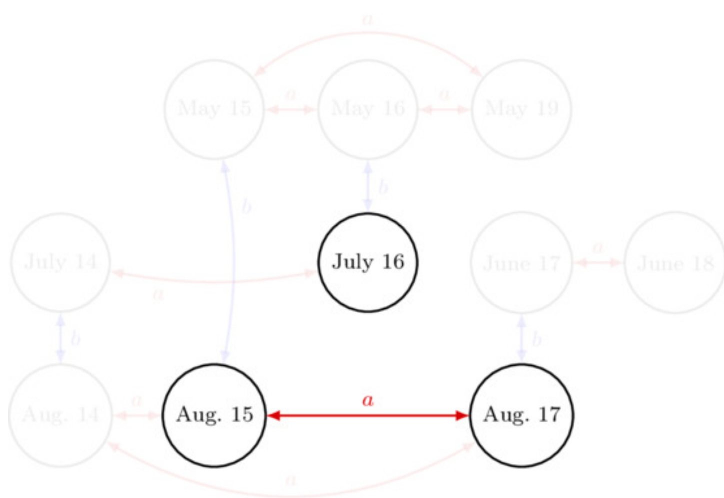
Now let us consider the effect of Bernard's first announcement, which we label as B_1:

(B_1) Bernard: I didn't know [Cheryl's birthday] originally, but now I do.

The first part of this statement ("I didn't know [Cheryl's birthday] originally") is to be interpreted in the model $M_{cb}[A_1]$ of Figure B2; however, this part uses the past tense ("did not know"), which suggests the interpretation must involve looking back at what was true in the original

model M_{cb} from Figure B1. The most straightforward way to do this is as follows: at date D in $M_{cb}[A_1]$, to determine if it is true that "Bernard did not know [Cheryl's birthday] originally", we look back to the same date in the model M_{cb} of Figure B1 and see if Bernard does know the birthday. If he does not, then we say that the statement holds in date D of model $M_{cb}[A_1]$; otherwise, we say the statement does not hold at date D of $M_{cb}[A_1]$. In this way, we determine whether to eliminate date D from $M_{cb}[A_1]$ in terms of what is true of that same date in the original model M_{cb}, which provides our straightforward interpretation of the past tense. Proceeding with this interpretation, we find that this first part of Bernard's announcement holds at every date in $M_{cb}[A_1]$: for each of these dates, if we look back to that same date in Figure B1, we see that Bernard does not know the birthday there (because there is a blue b-arrow pointing to a different date in Figure B1). Therefore the first part of this announcement does not cause us to eliminate any worlds.

How about the second part ("now I do [know Cheryl's birthday]")? For this, we must eliminate any date in $M_{cb}[A_1]$ of Figure B2 that makes the statement false: these are the dates in $M_{cb}[A_1]$ at which Bernard does *not* know the birthday, which are the dates in $M_{cb}[A_1]$ that have a blue b-arrow pointing to a different date. Examining Figure B2, it is clear that there are only two such dates: July 14 and August 14. Therefore, the net result of Bernard's first announcement B_1 is to eliminate these worlds. This gives us the model $M_{cb}[A_1][B_1]$ pictured in Figure B3.



$M_{cb}[A_1][B_1]$

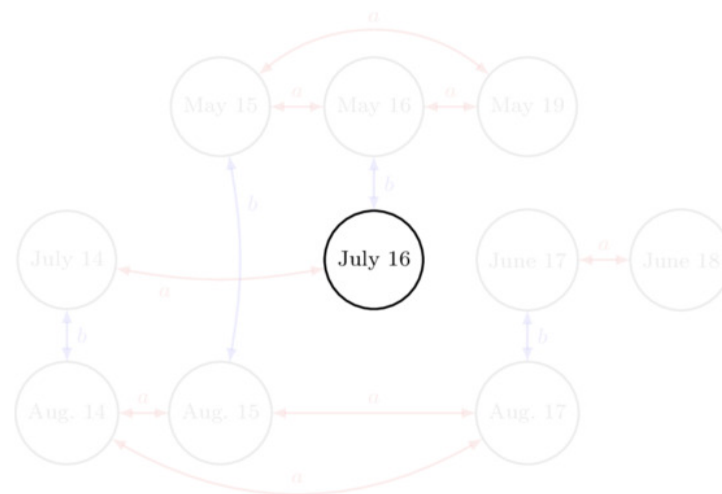
Figure B3. The model $M_{cb}[A_1][B_1]$ obtained from $M_{cb}[A_1]$ (Figure B2) after Bernard's first public announcement B_1. Eliminated worlds and arrows from the models M_{cb} and $M_{cb}[A_1]$ are faintly drawn in this picture for clarity; these are not actually present in the model. Also, reflexive arrows for each agent at each world have been omitted for simplicity.

Now let us look at the effect of Albert's second announcement, which we label as A_2:

(A_2) Albert: Well, now I know [Cheryl's birthday] too!

This public announcement has us eliminate from $M_{cb}[A_1][B_1]$ of Figure B3 all dates at which the statement is false: these are the dates in $M_{cb}[A_1][B_1]$ at which Albert does *not* know the birthday, which are the dates in $M_{cb}[A_1][B_1]$ at which there is a red a-arrow pointing to a different date. Examining Figure B3, it is clear that there are only two such dates: August 15 and August 17. Therefore, the result of Albert's second

announcement A_2 is to eliminate these worlds. This gives us the model $M_{cb}[A_1][B_1][A_2]$ pictured in Figure B4.



$M_{cb}[A_1][B_1][A_2]$

Figure B4. The model $M_{cb}[A_1][B_1][A_2]$ obtained from $M_{cb}[A_1][B_1]$ (Figure B3) after Albert's second public announcement A_2. Eliminated worlds and arrows from the models M_{cb}, $M_{cb}[A_1]$, and $M_{cb}[A_1][B_1]$ are faintly drawn in this picture for clarity; these are not actually present in the model. Also, reflexive arrows for each agent at each world have been omitted for simplicity.

So we obtain the solution: the only possible date that is consistent with the information provided in the problem statement is July 16.

1.3 Formalizing the Solution

The correctness of our solution depends on the assumption that the sequence of models M_{cb}, $M_{cb}[A_1]$, $M_{cb}[A_1][B_1]$, and $M_{cb}[A_1][B_1][A_2]$ we have identified is the only possible sequence that is consistent with the problem statement. Since we have not proved that this is so, we cannot be said to have proved that the solution is correct. Therefore, while our picture-based "solution" is visually appealing, it is not mathematically complete because it does not provide a proof.

To prove that July 16 is the correct answer, we must develop a mathematical formalization of the problem and then prove that this formalization implies that July 16 is the answer. Then the correctness of our solution rests only on the correctness of our formalization itself: so long as the formalization is all right, the solution must be correct. Note: one might argue that our picture-based solution is one such formalization; however, we have not proved that the starting model M_{cb} is the only Kripke model that adequately represents the initial setup, nor have we proved that our interpretation of which worlds to eliminate is correct! However, we will see that our picture-based solution *can* be formalized.

In this section, we consider two ways of filling this gap, one syntactic (i.e., based on proofs) and one semantic (i.e., based on models). The semantic formalization provides the missing argument that our picture-based solution works, whereas the syntactic formalization uses the theory $\mathsf{PAL}{+}C$ to prove that the solution we identified above is correct.

A proper formalization would use a formal language that can capture the past-tense part of Bernard's first announcement B_1 ("I didn't know [Cheryl's birthday] originally"). However, since we saw that the past-tense part provided no additional information (in that it eliminated no dates), we will simplify the problem by ignoring this part of Bernard's announcement, so that we may define a formal language that does not address tense issues. (The interested reader is directed to Appendix N for details on some approaches one might use in order to incorporate tense into the formalization and give a "true" formalization of the original problem.) So strictly speaking, our formalization here is for a variation of

the original problem in which Bernard's announcement is changed by eliminating the past-tense part; we might call this the "tense-free version" of the problem. (Note that we expect the formalization of both versions to lead to the same solution, though we have not proved it!)

1.3.1 Syntactic Formalization

Our set $\mathcal{A}$ of agents will be "a" for Albert and "b" for Bernard:

$$\mathcal{A} := \{a, b\}.$$

The set $\mathcal{P}$ of propositional letters of our language will be the dates:

$$\mathcal{P} := \left\{\begin{array}{l}\text{May 15, May 16, May 19,}\\ \text{June 17, June 18,}\\ \text{July 14, July 16,}\\ \text{Aug. 14, Aug. 15, Aug. 17}\end{array}\right\}.$$

For convenience, we define the following subsets of $\mathcal{P}$:

$$\begin{aligned}
\mathsf{May} &:= \{\text{May 15, May 16, May 19}\},\\
\mathsf{June} &:= \{\text{June 17, June 18}\},\\
\mathsf{July} &:= \{\text{July 14, July 16}\},\\
\mathsf{Aug} &:= \{\text{Aug. 14, Aug. 15, Aug. 17}\},\\
\mathsf{14} &:= \{\text{July 14, Aug. 14}\},\\
\mathsf{15} &:= \{\text{May 15, Aug. 15}\},\\
\mathsf{16} &:= \{\text{May 16, July 16}\},\\
\mathsf{17} &:= \{\text{June 17, Aug. 17}\},\\
\mathsf{18} &:= \{\text{June 18}\},\\
\mathsf{19} &:= \{\text{May 19}\}.
\end{aligned}$$

We also define the following sets of subsets of $\mathcal{P}$:

$$\mathsf{month} := \{\mathsf{May}, \mathsf{June}, \mathsf{July}, \mathsf{Aug}\},$$
$$\mathsf{day} := \{14, 15, 16, 17, 18, 19\}.$$

We use the sets $\mathcal{A}$ and $\mathcal{P}$ to define the language (PAL+C) of Public Announcement Logic with common knowledge defined in the main article:

$$F ::= p \mid F \wedge F \mid \neg F \mid [a]F \mid [F!]F \mid [B*]F \qquad \text{(PAL+C)}$$
$$p \in \mathcal{P},\ a \in \mathcal{A},\ B \subseteq \mathcal{A}$$

The information Cheryl provides establishes the following assumptions:

1. **It is common knowledge that Cheryl's birthday is one of those in $\mathcal{P}$.** We represent this using the formula

$$\mathsf{A}_1 := [\mathcal{A}*]\left(\bigwedge_{d \in \mathcal{P}} \langle \mathcal{A}* \rangle d\right) \wedge \left([\mathcal{A}*] \bigvee \mathcal{P}\right),$$

which says: it is common knowledge that for each of the dates d in the set $\mathcal{P}$, the agents jointly entertain the possibility that Cheryl's birthday might be d; further, it is common knowledge that Cheryl's birthday must be one of those in $\mathcal{P}$. This assumption ensures that Albert and Bernard do not exclude any of the dates from Cheryl's initial list and that the only dates they consider are those on this list.

2. **It is common knowledge that each of the dates in $\mathcal{P}$ is different.** We represent this using the formula

$$\mathsf{A}_2 := [\mathcal{A}*] \bigwedge_{d \in \mathcal{P}} \left(d \to \bigwedge_{d' \in \mathcal{P} - \{d\}} \neg d' \right),$$

which says: it is common knowledge that each date d in the set $\mathcal{P}$ is different than any other date d' in $\mathcal{P}$. This "obvious" assumption is required so that Albert and Bernard do not mistakenly believe different dates can coincide.

3. **It is common knowledge that Albert knows the month of Cheryl's birthday.** We represent this using the formula

$$\mathsf{A}_3 := \bigwedge_{m \in \mathsf{month}} [\mathcal{A}*] \left(\bigvee m \to \left([a] \left(\bigvee m \right) \wedge \bigwedge_{d \in m} \langle a \rangle d \right) \right),$$

which says: for each month m appearing in Cheryl's list, it is common knowledge that if the month of Cheryl's birthday is m, then Albert knows this, although he considers each of the dates d in m to be possible. This assumption ensures that Albert knows the month of Cheryl's birthday but that his knowledge is incomplete because he cannot exclude any date from the list that has the same month as the birthday.

4. **It is common knowledge that Bernard knows the day of Cheryl's birthday.** We represent this using the formula

$$\mathsf{A}_4 := \bigwedge_{d \in \mathsf{day}} [\mathcal{A}*] \left(\bigvee d \to \left([b] \left(\bigvee d \right) \wedge \bigwedge_{d' \in d} \langle a \rangle d' \right) \right),$$

which says: for each numerical day d appearing in Cheryl's list, it is common knowledge that if the day of Cheryl's birthday is d, then Bernard knows this, although he considers each of the dates d' having the same numerical day as d to be possible. This assumption ensures that Bernard knows the day of Cheryl's birthday but that his knowledge is incomplete because he cannot exclude any date from the list that has the same numerical day as the birthday.

We write "Setup" to denote the conjunction of the assumptions:

$$\mathsf{Setup} := \mathsf{A}_1 \wedge \mathsf{A}_2 \wedge \mathsf{A}_3 \wedge \mathsf{A}_4 .$$

For an agent $x \in \mathcal{A}$ to know Cheryl's birthday, he must only consider one date to be possible. We express this using the formula

$$\mathrm{KB}_x := \bigwedge_{d \in \mathcal{P}} (d \to [x]d),$$

which says: for each of the dates d in Cheryl's list, if d is her birthday, then agent x knows this. We use this formula to formalize the public announcements from Albert and Bernard's conversation, listed here in order:

1. $\neg\mathrm{KB}_a \wedge [a]\neg\mathrm{KB}_b$ — Albert does not know the birthday, and he knows that Bernard does not either;
2. KB_b — Bernard knows the birthday; and
3. KB_a — Albert knows the birthday.

Taken together, to verify our solution, we need to derive in the logic $\mathsf{PAL}{+}C$ the formula:

$$\mathsf{Setup} \to [\neg\mathrm{KB}_a \wedge [a]\neg\mathrm{KB}_b!][\mathrm{KB}_b!][\mathrm{KB}_a!](\text{July } 16). \tag{CB}$$

This formula says: given the initial setup, after Albert and Bernard have their conversation, the only possible date is July 16.

To complete the formalization, can the reader derive (CB) in $\mathsf{PAL}{+}C$?

1.3.2 Semantic Formalization

The semantic formalization takes our picture-based solution and fills in the missing details so as to prove that the sequence of models we used to derive an answer indeed gives us the correct answer. The basic idea of this solution is the following: though we cannot use the language (PAL+C) to capture the exact models we use in our picture-based solution, we can capture these models *up to bisimulation*. (The reader is directed to

Appendix G for the definition of bisimulation, the statement of the Bisimulation Theorem, and the notation that we define in each of these.)

To begin, we define our sets $\mathcal{A}$ of agents and $\mathcal{P}$ of propositional letters just as we did for the syntactic solution. Likewise, we define the formulas A_1 through A_4, Setup, KB_a, and KB_b as we did before.

We shall restrict attention to the class $\mathcal{C}$ of Kripke models M whose agent set is $\mathcal{A}$ and whose set of propositional letters is $\mathcal{P}$. This means that $M \in \mathcal{C}$ if and only if M contains a binary accessibility relation $R^M_x \subseteq W^M \times W^M$ for each of our agents $x \in \mathcal{A}$ and no others and the valuation V^M in M has functional type $V^M : \mathcal{P} \to W^M$ (i.e., this valuation maps letters in the set $\mathcal{P}$, and not some other set, to sets of worlds in M). We let $\mathcal{C}_*$ be the class of pointed Kripke models (M, w) having $M \in \mathcal{C}$.

Given a Kripke model $M \in \mathcal{C}$, a world $w \in W^M$, and an agent $x \in \mathcal{A}$, we define the *connected component*

$$\mathsf{cc}^M(w) := \left\{ v \in W^M \mid w\left(R^M_{\mathcal{A}}\right)^* v \right\}$$

of w. This is the set of all worlds v in M that can be reached from w by following zero or more agent arrows. (And hence $\mathsf{cc}^M(w)$ includes w itself.) The idea is that $\mathsf{cc}^M(w)$ is the set of all worlds that the agents would jointly consider possible from a "common knowledge" perspective. (In Appendix A, we define $R^M_{\mathcal{A}} := \bigcup_{x \in \mathcal{A}} R^M_x$ and use R^* to denote the application of the reflexive-transitive closure operator to the binary relation R.)

It shall now be our task to prove the following theorem.

Setup Theorem. Given $(M_1, w_1) \in \mathcal{C}_*$ and $(M_2, w_2) \in \mathcal{C}_*$, if there exists $d_* \in \mathcal{P}$ such that

$$M_1, w_1 \models d_* \wedge \mathsf{Setup} \text{ and } M_2, w_2 \models d_* \wedge \mathsf{Setup},$$

then $(M_1, w_1) \leftrightarrows (M_2, w_2)$.

Proof. Assume there exists $d_* \in \mathcal{P}$ such that $M_1, w_1 \vDash d_* \wedge \mathsf{Setup}$ and $M_2, w_2 \vDash d_* \wedge \mathsf{Setup}$. It follows that the pointed Kripke models (M_1, w_1) and (M_2, w_2) both satisfy the assumption formulas A_1 through A_4. In particular, since they satisfy A_1 and A_2, it follows by inspection of these formulas and the definition of satisfaction that for each $i \in \{1,2\}$, there exists a surjection $f_i : \mathsf{cc}^{M_i}(w_i) \to \mathcal{P}$ satisfying the property that for each world $u \in \mathsf{cc}^{M_i}(w_i)$ and date $d \in \mathcal{P}$, we have $M_i, u \vDash d$ if and only if $f_i(u) = d$. Intuitively, this says that each world in the connected component $\mathsf{cc}^{M_1}(w_1)$ corresponds to a unique date from Cheryl's list; the same goes for worlds in the connected component $\mathsf{cc}^{M_2}(w_2)$. Define the inverse $f_i^{-1} : \mathcal{P} \to \wp(\mathsf{cc}^{M_i}(w_i))$ by setting $f_i^{-1}(d) := \{u \in \mathsf{cc}^{M_i}(w_i) \mid f_i(u) = d\}$. So $f_i^{-1}(d)$ is the set of worlds in $\mathsf{cc}^{M_i}(w_i)$ that are assigned to date d. We define the following binary relations between W^{M_1} and W^{M_2}:

$$
\begin{aligned}
Z_0 &:= \{(w_1, w_2)\}, \\
Z_{i+1} &:= \{(u_1', u_2') \mid f_1(u_1') = f_2(u_2') \wedge \exists x \in \mathcal{A}, u_1, u_2 : u_1 Z_i u_2 \\
&\qquad\qquad \wedge u_1 R_x^{M_1} u_1' \wedge u_2 R_x^{M_2} u_2'\}, \\
Z &:= \textstyle\bigcup_{i=0}^{\infty} Z_i.
\end{aligned}
$$

Since $w_1 Z w_2$, it follows that Z is nonempty. By induction on i, one can prove that $u_1 Z_i u_2$ implies $f_1(u_1) = f_2(u_2)$. The base case, where $i = 0$, follows because because (M_1, w_1) and (M_2, w_2) both satisfy the same letter $d_* \in \mathcal{P}$ (and no others). For the induction step, where $i > 0$, the result follows by the definition of Z_i. Therefore, $u_1 Z u_2$ implies $f_1(u_1) = f_2(u_2)$. Making use of this fact, and referring now to the definition of bisimulation (Appendix G), it follows by the definitions of f_1 and f_2 that Z satisfies Propositional Identity. It shall now be our task to prove that Z also satisfies Back. For this, we first note that $u_1 Z_i u_2$ implies $u_1 \in \mathsf{cc}^{M_1}(w_1)$ and $u_2 \in \mathsf{cc}^{M_2}(w_2)$; this is proved by a straightforward induction on i. To prove Back, we must show that if $u_1 Z u_2$ and $u_2 R_x^{M_2} u_2'$ for some $x \in \mathcal{A}$, then there exists $u_1' \in W^{M_1}$ such that $u_1 R_x^{M_1} u_1'$ and $u_1' Z u_2'$. Since $\mathcal{A} = \{a, b\}$, we may break up this proof into two items, one for $x = a$ and one for $x = b$.

- If $u_1 Z u_2$ and $u_2 R_a^{M_2} u_2'$, then there exists $u_1' \in W^{M_1}$ such that $u_1 R_a^{M_1} u_1'$ and $u_1' Z u_2'$.

  Here we will actually prove the following for each $i \geq 0$: if $u_1 Z_i u_2$ and $u_2 R_a^{M_2} u_2'$, then there exists $u_1' \in W^{M_1}$ such that $u_1 R_a^{M_1} u_1'$ and $u_1' Z_{i+1} u_2'$. Since we have $u_1 Z u_2$ if and only if there exists k such that $u_1 Z_k u_2$, it follows that what we will actually prove implies the desired result. So let us proceed with our proof. Assume $u_1 Z_i u_2$ and $u_2 R_a^{M_2} u_2'$. It follows that $f_1(u_1) = f_2(u_2)$, $u_1 \in \mathsf{cc}^{M_1}(w_1)$, and $u_2 \in \mathsf{cc}^{M_2}(w_2)$. Since our collection month (of "months") consists of pairwise disjoint sets of dates and every date in $\mathcal{P}$ occurs in exactly one set in the collection, it follows from $M_2, w_2 \vDash \mathsf{A}_1 \wedge \mathsf{A}_2 \wedge \mathsf{A}_3$ and $u_2 \in \mathsf{cc}^{M_2}(w_2)$ that we have $M_2, u_2 \vDash \bigvee m$ for some $m \in \mathsf{month}$. Thus $f_2(u_2) \in m$. In words: date $f_2(u_2)$ falls in month m. Further, since $M_2, w_2 \vDash \mathsf{A}_3$ and $u_2 \in \mathsf{cc}^{M_2}(w_2)$, it follows that $M_2, u_2 \vDash [a](\bigvee m)$. Therefore, since $u_2 R_a^{M_2} u_2'$, we have $M_2, u_2' \vDash \bigvee m$. Hence $f_2(u_2') \in m$, so that date $f_2(u_2')$ falls in month m as well. Since $f_1(u_1) = f_2(u_2)$, the date $f_1(u_1)$ also falls in month m, and so $M_1, u_1 \vDash \bigvee m$. Applying the fact that $M_1, w_1 \vDash \mathsf{A}_3$ and $u_1 \in \mathsf{cc}^{M_1}(w_1)$, it follows by $M_1, u_1 \vDash \bigvee m$ and the definition of A_3 that $M_1, u_1 \vDash \bigwedge_{d \in m} \langle a \rangle d$. Since $f_2(u_2') \in m$, we obtain $M_1, u_1 \vDash \langle a \rangle f_2(u_2')$. As a consequence of this, there exists $u_1' \in W^{M_1}$ such that $u_1 R_a^{M_1} u_1'$ and $M_1, u_1' \vDash f_2(u_2')$. Hence $f_1(u_1') = f_2(u_2')$. But then we have shown that $f_1(u_1') = f_2(u_2')$, $u_1 Z_i u_2$, $u_1 R_a^{M_1} u_1'$, and $u_2 R_a^{M_2} u_2'$. Applying the definition of Z_{i+1}, it follows that $u_1' Z_{i+1} u_2'$.

- If $u_1 Z u_2$ and $u_2 R_b^{M_2} u_2'$, then there exists $u_1' \in W^{M_1}$ such that $u_1 R_b^{M_1} u_1'$ and $u_1' Z u_2'$.

  The argument is almost the same as for the previous item, except that we use b instead of a, use A_4 instead of A_3, use day instead of month, and refer to days instead of months.

So Z satisfies Back. By a similar argument (that interchanges 1's and 2's in the appropriate places), it follows that Z satisfies Forth as well. So Z is a bisimulation between M_1 and M_2. To complete the proof, we observe that since $w_1 Z_0 w_2$, we obtain $w_1 Z w_2$. Conclusion: $(M_1, w_1) \leftrightarrows (M_2, w_2)$. ■

We will also make use of the following result about preservation of bisimulation under public announcements by (ML)-formulas:

Bisimulation Preservation Theorem. Let (M_1, w_1) and (M_2, w_2) be pointed Kripke models and F be an (ML)-formula such that $M_i, w_i \vDash F$ for each $i \in \{1,2\}$. If $(M_1, w_1) \leftrightarrows (M_2, w_2)$, then $(M_1[F], w_1) \leftrightarrows (M_2[F], w_2)$.

Proof. Let Z be the bisimulation between M_1 and M_2 satisfying $w_1 Z w_2$. We prove that

$$Z' := Z \cap (W^{M_1}[F] \times W^{M_2}[F])$$

is a bisimulation between $M_1[F]$ and $M_2[F]$ satisfying $w_1 Z' w_2$. Proceeding, Propositional Identity for Z' follows by the definition of Z' and Propositional Identity for Z. To prove Back for Z', assume we have $u_1 Z' u_2$ and $u_2 R_x^{M_2[F]} u_2'$. It follows that $u_1 Z u_2$, $u_2 R_x^{M_2} u_2'$, and $M_2, u_2' \vDash F$. Applying the Back property for Z, there exists $u_1' \in W^{M_1}$ such that $u_1 R_x^{M_1} u_1'$ and $u_1' Z u_2'$. But then $(M_1, u_1') \leftrightarrows (M_2, u_2')$, so since we have $F \in$ (ML) and $M_2, u_2' \vDash F$, it follows by the Bisimulation Theorem that $M_1, u_1' \vDash F$ and therefore $u_1' \in W^{M_1[F]}$. Hence $u_1 R_x^{M_1[F]} u_1'$ and $u_1' Z' u_2'$. This completes the proof of Back for Z'. Forth for Z' is proved similarly. Finally, we have $w_1 Z' w_2$ by the assumption that $M_i, w_i \vDash F$ for each $i \in \{1,2\}$ and the definition of Z'. ■

Referring to the model M_{cb} from our picture-based solution (Figure B1), it is not difficult to verify that we have each of the following:

- M_{cb} has $\mathcal{A}$ as its agent set and $\mathcal{P}$ as its set of propositional letters; that is, $M_{cb} \in \mathcal{C}$.
- We have $M_{cb}, w \vDash \mathsf{Setup}$ for each world $w \in W^{M_{cb}}$.
- Each world w in M_{mc} satisfies exactly one date $d_w \in \mathcal{P}$, meaning we have $M_{mc}, w \vDash d_w$ and $M_{mc}, w \nvDash d'$ for each $d' \in \mathcal{P} - \{d_w\}$.

Choose an arbitrary $M \in \mathcal{C}$ satisfying the property that $M, v \vDash \mathsf{Setup}$ for each world $v \in W^M$. Since $M \in \mathcal{C}$ (and so M is based on the propositional letter set $\mathcal{P}$), each world w in M_{cb} gives rise to a set

$$g(w) := \{v \in W^M \mid M, v \vDash d_w\}$$

of worlds in M that also satisfy the same date d_w. This defines a function $g : W^{M_{cb}} \to \wp(W^M)$ mapping worlds in M_{cb} to sets of worlds in M. Since every world in M satisfies Setup, it is not difficult to see that $g(w) \neq \emptyset$ for each w in M_{mc}. Further, since worlds in M all satisfy Setup and since each date in $\mathcal{P}$ is true at a unique world in M_{mc}, it follows that every world $v \in W^M$ gives rise to a unique world $g^{-1}(v) \in W^{M_{mc}}$ such that $g(g^{-1}(v)) = v$. Given w in M_{cb}, since w and the worlds in $g(w)$ all satisfy $d_w \wedge \mathsf{Setup}$, it follows by the Setup Theorem that $(M_{cb}, w) \leftrightarrows (M, v)$ for each $v \in g(w)$. Therefore, for each sequence $F_1, \dots, F_n$ of (ML)-formulas such that the structures $M_{cb}[F_1] \cdots [F_n]$ and $M[F_1] \cdots [F_n]$ contain nonempty sets of worlds (and therefore these structures are Kripke models), it follows by the Bisimulation Preservation Theorem that

$$(M_{cb}[F_1]\cdots[F_n],w) \leftrightarrows (M[F_1]\cdots[F_n],v) \text{ for each } v \in g(w).$$

For a Kripke model $M \in \mathcal{C}$ and a date $d \in \mathcal{P}$, we shall say that a sequence $F_1,\dots,F_n$ of (ML)-formulas *yields date d in M* to mean that the structure $M[F_1]\cdots[F_n]$ contains worlds (so this is a Kripke model) and each of these worlds satisfies a unique date $d \in \mathcal{P}$. This leads us to the following result:

Agreement Theorem. Suppose $M \in \mathcal{C}$ satisfies $M,w \models \mathsf{Setup}$ for each world $w \in W^M$. The sequence $F_1,\dots,F_n$ of (ML)-formulas yields date d in M_{mc} if and only if this same sequence yields date d in M.

Proof. This proof uses the notation and results established just above the statement of the theorem. Since each date in $\mathcal{P}$ is satisfied at a unique world in M_{cb}, it follows that $F_1,\dots,F_n$ yields date d in M_{mc} if and only if $M_{mc}[F_1]\cdots[F_n]$ contains a unique world w such that $d_w = d$. We make use of this in what follows. Proceeding, assume that $F_1,\dots,F_n$ yields date d in M_{cb}, so that w is the unique world in $M_{mc}[F_1]\cdots[F_n]$ and $d = d_w$. It follows that for each world $u \in W^{M_{cb}} - \{w\}$, there exists a minimum $k(u) \in \{1,\dots,n-1\}$ such that $M_{cb}[F_1]\cdots[F_{k(u)-1}],u \not\models F_{k(u)}$. Now we have seen that $g(w) \neq \emptyset$. Further, since we have for each $i \in \{1,\dots,n\}$ and $v \in g(w)$ that

$$(M_{cb}[F_1]\cdots[F_i],w) \leftrightarrows (M[F_1]\cdots[F_i],v),$$

it follows by the Bisimulation Theorem that $M[F_1]\cdots[F_n]$ contains the worlds in $g(w)$. To see that it contains no other worlds, notice that for any $v \in W^M - g(w)$, we have that

$$M_{cb}[F_1]\cdots[F_{k(g^{-1}(v))-1}],g^{-1}(v) \not\models F_{k(g^{-1}(v))}$$

and therefore, since we have a bisimulation linking $g^{-1}(v)$ in $M_{cb}[F_1]\cdots[F_{k(g^{-1}(v))-1}]$ to v in $M[F_1]\cdots[F_{k(g^{-1}(v))-1}]$, it follows by the Bisimulation Theorem that

$$M[F_1]\cdots[F_{k(g^{-1}(v))-1}],v \not\models F_{k(g^{-1}(v))}.$$

As a result, v is not in $M[F_1]\cdots[F_n]$. Since $v \in W^M - g(w)$ was chosen arbitrarily, we have proved that $F_1,\dots,F_n$ yields date $d_w = d$ in M. This proves one direction of the theorem. For the converse, we assume that $F_1,\dots,F_n$ yields date d in M. It follows that $M[F_1]\cdots[F_n]$ consists of a nonempty set W of worlds and for each $v \in W$, we have $M[F_1]\cdots[F_n],v \models d$. Further, for each world $u \in W^M - W$, there exists a minimum $k(u) \in \{1,\dots,n-1\}$ such that $M[F_1]\cdots[F_{k(u)-1}],u \not\models F_{k(u)}$. Since there is a bisimulation linking each v in M to $g^{-1}(v)$ in M_{cb}, it follows by the Bisimulation Preservation Theorem that $M_{cb}[F_1]\cdots[F_n]$ has $\{g(v) \mid v \in W\}$ as a subset of its set of worlds. However, since we assumed that $F_1,\dots,F_n$ yields date d in M and since M_{mc} has exactly one world satisfying a given date, it follows $\{g(v) \mid v \in W\}$ is a singleton $\{w\}$. To show that $M_{cb}[F_1]\cdots[F_n]$ contains no worlds other than this w, take an arbitrary $u \in W^{M_{cb}} - \{w\}$. Since

$$M[F_1]\cdots[F_{k(g(u))-1}],g(u) \not\models F_{k(g(u))}$$

and we have a bisimulation linking $g(u)$ in $M[F_1]\cdots[F_{k(g(u))-1}]$ to u in $M_{cb}[F_1]\cdots[F_{k(g(u))-1}]$, it follows by the Bisimulation Theorem that

$$M_{cb}[F_1]\cdots[F_{k(g(u))-1}],u \not\models F_{k(g(u))}$$

and therefore u is not in $M[F_1]\cdots[F_n]$. As a result, we may conclude that the set of worlds of the latter model is $\{w\}$ and therefore $F_1,\dots,F_n$ yields date d in M_{mc} as well. ∎

We shall say that a picture-based solution to the Cheryl's birthday problem is *legitimate* if and only if it satisfies each of the following items:

- The solution starts with a Kripke model M based on agent set $\mathcal{A}$ and propositional letter set $\mathcal{P}$; that is, we have $M \in \mathcal{C}$.
- Each world of the starting model M satisfies the initial assumptions of the problem; that is, we have $M, w \vDash \mathsf{Setup}$ for each world $w \in W^M$.
- The solution produces date $d \in \mathcal{P}$ as its answer if and only if the sequence

$$\neg \mathrm{KB}_a \wedge [a]\neg \mathrm{KB}_b,\ \mathrm{KB}_b,\ \mathrm{KB}_a \qquad \text{(CBS)}$$

  of (ML)-formulas yields date d in M.

This completes the definition of "legitimate" picture-based solutions. Note that this definition omits the condition that the starting model be finite. One might wish to add this condition so as to guarantee that the picture-based solution can actually be drawn using the notation for Kripke models from Appendix A. For those so inclined, this further condition may be added without affecting the remainder of our argument. However, in the interest of having "legitimate" picture-based solutions be the largest class possible, we do not add this requirement here.

We use the Agreement Theorem to obtain the following result:

> **Picture-based Solution Theorem.** Every legitimate picture-based solution to the Cheryl's Birthday Problem produces the same answer.
>
> *Proof*. Our picture-based solution from above is legitimate: we have $M_{cb} \in \mathcal{C}$, every world of M_{cb} satisfies Setup, and we produced our answer by showing that the sequence (CBS) yields July 16 in M_{cb}. Applying the Agreement Theorem, every other legitimate picture-based solution must also produce July 16 as its answer. ■

This theorem tells us why our picture-based solution from above is correct: no legitimate picture-based solution can disagree with the answer we obtained.

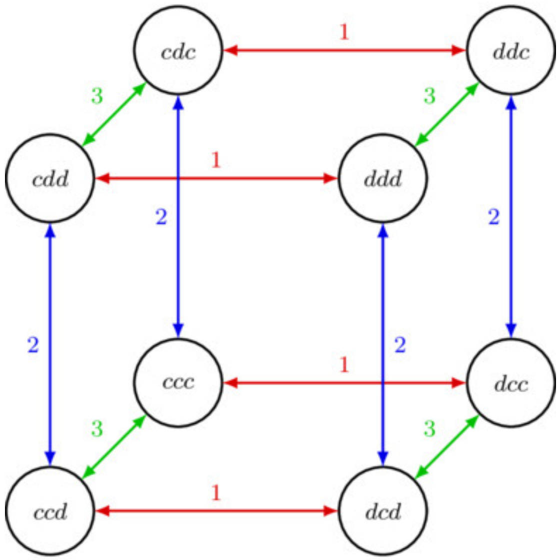
2. The Muddy Children Puzzle

2.1 Problem Statement

> **The Muddy Children Puzzle.** Three children are playing in the mud. Father calls the children to the house, arranging them in a semicircle so that each child can clearly see every other child. "At least one of you has mud on your forehead", says Father. The children look around, each examining every other child's forehead. Of course, no child can examine his or her own. Father continues, "If you know whether your forehead is dirty, then step forward now". No child steps forward. Father repeats himself a second time, "If you know whether your forehead is dirty, then step forward now". Some but not all of the children step forward. Father repeats himself a third time, "If you know whether your forehead is dirty, then step forward now". All of the remaining children step forward. How many children have muddy foreheads?

2.2 Solution

Each of the three children has a forehead that is either dirty or clean. Using c for "clean" and d for "dirty", a sequence of three letters may be used to indicate the state of affairs. For example, cdc indicates that the first child has a clean forehead, the second child has a dirty forehead, and the third child has a clean forehead. As a result, there are 8 possibilities in total. This leads to the model M_{mc} pictured in Figure B5.



M_{mc}

Figure B5. Initial setup for the Muddy Children Puzzle. Red 1-arrows indicate the first child's uncertainty, blue 2-arrows indicate the second child's uncertainty, and green 3-arrows indicate the third child's uncertainty. For simplicity, the picture omits reflexive arrows (i.e., to complete the picture, one must add for each child and each world an arrow for that child pointing in a loop from that world right back to that same world).

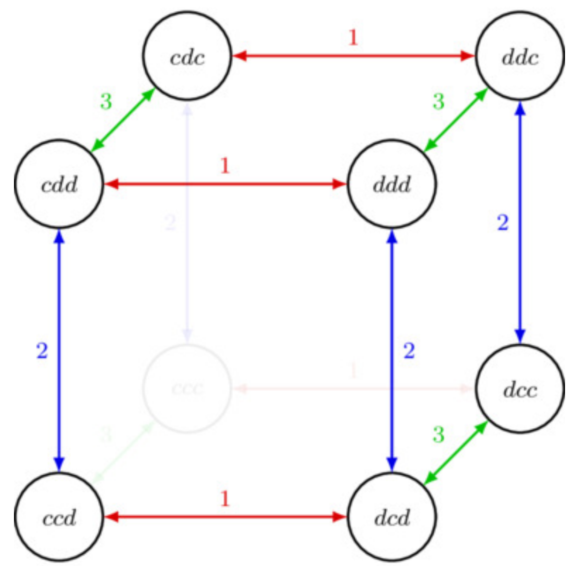
In the model M_{mc} of Figure B5, we see that each child knows the status of the other children's forehead but not that of his own: at each world, the only worlds a child considers possible are those in which the status of the other children's foreheads is the same but the status of his own forehead might be different. For example, at world *ddc*, the only possibilities the first child entertains are *ddc* itself (via the non-drawn reflexive 1-arrow) and the world *cdc*. Since the status of the other children's foreheads is the same at worlds *ddc* and *cdc* (i.e., 2's forehead is dirty and 3's is clean) but

the status of the first child's forehead is not the same (i.e., 1's forehead is dirty in *ddc* but clean in *cdc*), we can say the following: at world *ddc*, the first child knows whether the other children's foreheads are dirty but does not know whether his own is dirty. We can show that this is so at every world and for every child. Therefore, M_{mc} indeed models the initial situation of the Muddy Children Puzzle.

Now let us consider the effect of Father's first announcement, which we label by F:

(F) "At least one of you has mud on your forehead."

The effect of this public announcement is to eliminate all those worlds at which F is false; these are the worlds at which no child is dirty. There is only one such world: *ccc*. Therefore, the public announcement of F modifies the initial situation M_{mc} by producing the mode $M_{mc}[F]$ pictured in Figure B6.



$M_{mc}[F]$

Figure B6. The model $M_{mc}[F]$ obtained from M_{mc} (Figure B5) by the public announcement of F. Worlds and arrows from M_{mc} that have been deleted are drawn faintly; these are not actually present in the model $M_{mc}[F]$. Also, the picture omits reflexive arrows for each agent at each world.

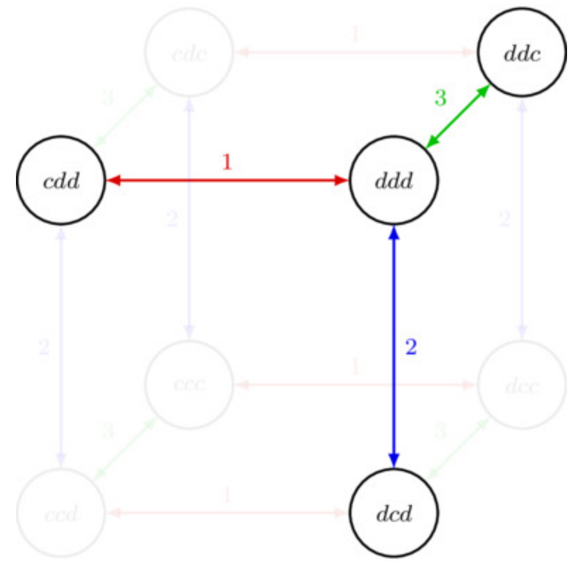
Father then says, “If you know whether your forehead is dirty, then step forward now”. This is the first time he says this and no child steps forward. We take Father’s statement followed by no child stepping forward to be equivalent to the following public announcement, which we denote by N:

(N) “No child knows whether his own forehead is dirty.”

The effect of an announcement of N is to eliminate those worlds at which N is false; these are the worlds at which some child knows whether his own forehead is dirty. For an agent to know whether his own forehead is dirty at a given world, he must know at that world that his forehead is dirty or he must know at that world that his forehead is clean. And we have that “agent a knows F at w” if and only if F is true at each world w' such that there is an a-arrow pointing from w to w' (of course taking into account non-drawn reflexive arrows). So, taken together, agent a knows whether his forehead is dirty just in case every outgoing a-arrow points only to worlds having the same clean/dirty status for agent a. Therefore, the effect of the public announcement of N is to eliminate worlds satisfying this property for one of the children. Referring to Figure B6, we see that the worlds to eliminate are:

- *cdc*, since this world has no 2-arrow pointing to a world with a different clean/dirty status for the second child;
- *ccd*, since this world has no 3-arrow pointing to a world with a different clean/dirty status for the third child; and
- *dcc*, since this world has no 1-arrow pointing to a world with a different clean/dirty status for the first child.

The result of the announcement of N is the model $M_{mc}[F][N]$ pictured in Figure B7.



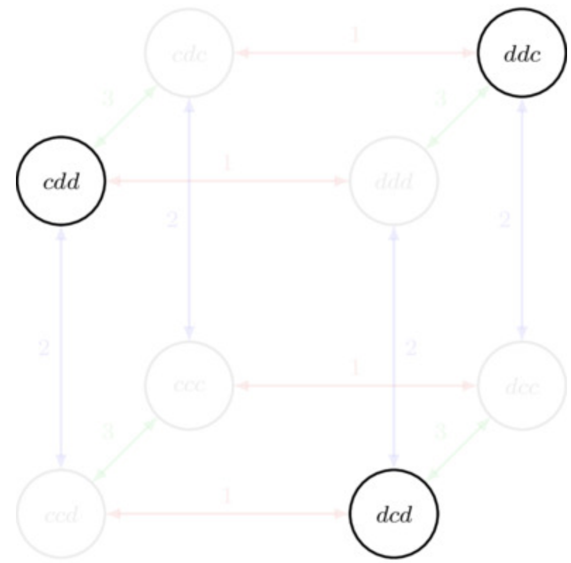
$M_{mc}[F][N]$

Figure B7. The model $M_{mc}[F][N]$ obtained from $M_{mc}[F]$ (Figure B6) after the public announcement of N. Worlds and arrows from M_{mc} and $M_{mc}[F]$ that have been deleted are drawn faintly; these are not actually present in the model $M_{mc}[F][N]$. Also, the picture omits reflexive arrows for each agent at each world.

Father now makes his announcement for the second time (“If you know whether your forehead is dirty, then step forward now”). But this time some children step forward, thereby conveying that these children know whether their foreheads are dirty. We therefore interpret Father's statement followed by the actions of the children as a single announcement of the negation $\neg N$ of our statement N from before. That is, we take the net result to be equivalent to the following public announcement:

($\neg N$) "Some child knows whether his forehead is dirty."

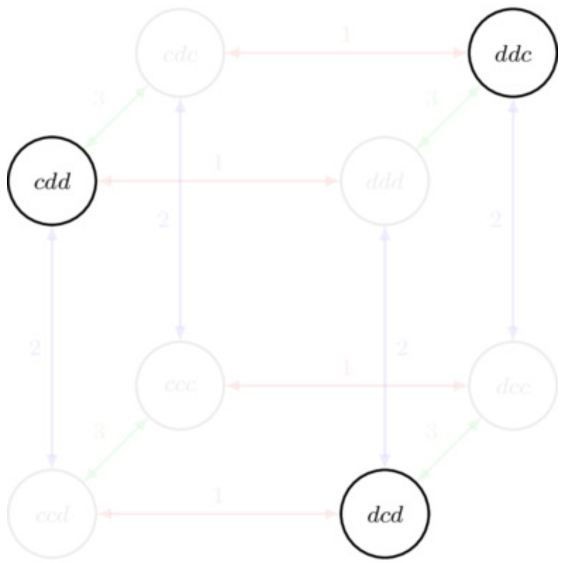
The effect of this announcement is to delete all those worlds at which the statement $\neg N$ is false: these are the worlds at which no child knows whether is forehead is dirty. That is, we are to delete worlds at which each agent has an arrow pointing to a world at which his forehead is clean and another arrow pointing to another world at which his forehead is dirty (of course taking into account non-drawn reflexive arrows). Examining Figure B7, we see that the only such world is *ddd*: at this world, every agent entertains a possibility at which his own forehead is clean along with a possibility at which his own forehead is dirty. Therefore, the result of the public announcement of $\neg N$ is to delete this world, thereby yielding the model $M_{mc}[F][N][\neg N]$ pictured in Figure B8.



$M_{mc}[F][N][\neg N]$

Figure B8. The model $M_{mc}[F][N][\neg N]$ obtained from $M_{mc}[F][N]$ (Figure B7) after the public announcement of $\neg N$. Worlds and arrows from M_{mc}, $M_{mc}[F]$, and $M_{mc}[F][N]$ that have been deleted are drawn faintly; these are not actually present in the model $M_{mc}[F][N][\neg N]$. Also, the picture omits reflexive arrows for each agent at each world.

Father now makes his announcement for the third time ("If you know whether your forehead is dirty, then step forward now"). Now, we are told, the remaining children step forward. We again take this to be equivalent to the announcement of $\neg N$ ("Some child knows whether his forehead is dirty"). As before, the effect of this announcement is to delete any worlds at which the announcement is false: these are the worlds at which no child knows whether his forehead is dirty, which are the worlds at which every child entertains a possibility at which his forehead is dirty and a possibility at which his forehead is clean. Examining Figure B8, we see that none of the worlds satisfy these requirements: at each world, each child only entertains that very world, and therefore each child knows whether his forehead is dirty. So the result of the second public announcement of $\neg N$ is the model $M_{mc}[F][N][\neg N][\neg N]$ pictured in Figure B9.



$M_{mc}[F][N][\neg N][\neg N]$

Figure B9. The model $M_{mc}[F][N][\neg N][\neg N]$ obtained from $M_{mc}[F][N]$ (Figure B8) after the public announcement of $\neg N$. Worlds and arrows from M_{mc}, $M_{mc}[F]$, and $M_{mc}[F][N]$ that have been deleted are drawn faintly; these are not actually present in the model $M_{mc}[F][N][\neg N][\neg N]$. Also, the picture omits reflexive arrows for each agent at each world.

Obviously, since the second announcement of $\neg N$ did not delete any worlds, the models of Figures B8 and B9 are the same:

$$M_{mc}[F][N][\neg N] = M_{mc}[F][N][\neg N][\neg N].$$

Looking now to this model, we obtain the solution: two children are dirty and one is clean.

2.3 Formalizing the Solution

Formalization of the Muddy Children Puzzle has certain similarities to the formalization of Cheryl's Birthday, though there are of course some differences. We refer the reader to Plaza (1989, 2007) for a full formalization of the Muddy Children Puzzle that uses public announcements. A similar solution that does not explicitly use this machinery may be found in Fagin et al. (1995).

3. The Sum and Least Common Multiple Puzzle

3.1 Problem Statement

The Sum and Least Common Multiple Puzzle. Referee reminds Mr. S and Mr. L that the least common multiple ("lcm") of two positive integers x and y is the smallest positive integer that is divisible without any remainder by both x and y (e.g., $\text{lcm}(3,6) = 6$ and $\text{lcm}(5,7) = 35$). Referee then says,

> Among the integers ranging from 2 to 7, including 2 and 7 themselves, I will choose two different numbers. I will whisper the sum to Mr. S and the least common multiple to Mr. L.

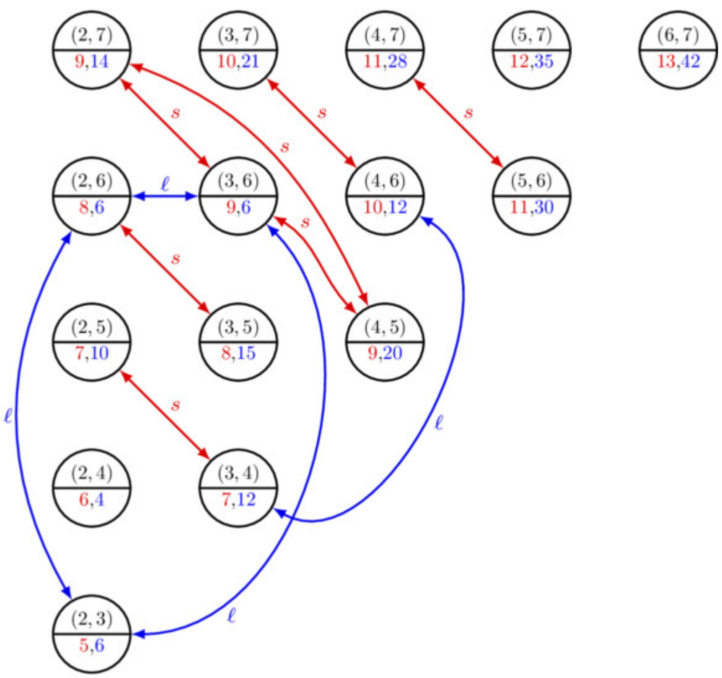
Referee then does as promised. The following dialogue then takes place:

> Mr. S: I know that you don't know the numbers.
> Mr. L: Ah, but now I do know them.
> Mr. S: And so do I!

What are the numbers?

3.2 Solution

Referee must choose two different integers in range $[2,7]$. Since the order they are chosen does not matter, there are fifteen ways for Referee to make this choice. For each choice, we calculate the sum and the least common multiple. If two choices have the same sum, then Mr. S should consider one possible relative to the other: he knows only the sum and so cannot distinguish between these two possibilities. If the two choices have the same least common multiple (henceforth "LCM"), then Mr. L should consider one possible relative to the other: he knows only the LCM and so cannot distinguish between these two possibilities. This leads to the model $M_{s\ell}$ pictured in Figure B10.



$M_{s\ell}$

Figure B10. The model $M_{s\ell}$ depicting the initial setup for the Sum and Least Common Multiple Puzzle. Each of the circles depicts a possible

choice by Referee: the pair of numbers in parenthesis above the horizontal line are the numbers chosen by Referee; below the line, the number on the left in red is the sum of the chosen numbers and the number on the right in blue is the least common multiple of the chosen numbers. Red arrows labeled by s represent Mr. S's uncertainties, and blue arrows labeled by ℓ represent Mr. L's uncertainties. For simplicity, reflexive arrows have been omitted (i.e., to complete the picture, one should add for each agent at each world an arrow for that agent pointing in a loop from the world in question back to that same world).

In Figure B10, we see that Mr. S has uncertainty between worlds having the same sum and Mr. L has uncertainty between worlds having the same LCM. Mr. S then makes his first statement, which we denote by S_1:

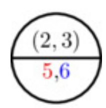
(S_1) Mr. S: I know that [Mr. L does not] know the numbers.

The effect of the public announcement of S_1 is to delete any world (i.e., possibility) at which S_1 is false: these are worlds at which Mr. S does not know that Mr. L does not know the numbers. Such worlds satisfy the property that there is a red s-arrow to a world at which Mr. L does know the numbers. And Mr. L knows the numbers at a given world if and only if there is no blue ℓ-arrow pointing to a different world. Taken together, the public announcement of S_1 should delete every world from which we can take a red s-arrow (possibly a reflexive one) to a world that has no ℓ-arrow pointing to a different world. Let us call such ℓ-arrows "outgoing". Examining Figure B10, we see that the following worlds are to be deleted:

- (2,7): from this world we can take a reflexive s-arrow to the world itself, which has no outgoing ℓ-arrows;
- (3,7): as for (2,7);
- (4,7): as for (2,7);
- (5,7): as for (2,7);

- (6,7): as for (2,7);
- (2,6): from this world we can take an s-arrow to (3,5), which has no outgoing ℓ-arrows;
- (3,6): from this world we can take an s-arrow to (2,7), which has no outgoing ℓ-arrows;
- (4,6): from this world we can take an s-arrow to (3,7), which has no outgoing ℓ-arrows;
- (5,6): as for (2,7);
- (2,5): as for (2,7);
- (3,5): as for (2,7);
- (4,5): as for (2,7);
- (2,4): as for (2,7); and
- (3,4): from this world we can take an s-arrow to (2,5), which has no outgoing ℓ-arrows.

The result is the model $M[S_1]$ pictured in Figure B11.



$M_{s\ell}[S_1]$

Figure B11. The model $M_{s\ell}[S_1]$ obtained from $M_{s\ell}$ (Figure B10) after the public announcement of S_1. The picture omits reflexive arrows for each agent at each world.

Note that in the model $M_{s\ell}[S_1]$, both agents know the numbers: each of the agents only considers one pair of numbers to be possible. From here we have Mr. L's first announcement, which we label as L:

(L) Mr. L: Ah, but now I do know [the numbers].

This announcement has the effect of removing any world at which L is false: these are the worlds at which Mr. L does not know the numbers. As

we have seen, no such worlds exist. Therefore the model $M_{s\ell}[S_1][L]$ obtained from $M_{s\ell}[S_1]$ by the public announcement of L is unchanged:

$$M_{s\ell}[S_1][L] = M_{s\ell}.$$

So Figure B11 also depicts $M_{s\ell}[S_1][L]$.

We then turn to Mr. S's second announcement, which we label as S_2:

(S_2) Mr. S: And now so do I [know the numbers]!

This announcement has the effect of removing any world at which S_2 is false: these are the worlds at which Mr. S does not know the numbers. But again we saw that there is no such world in the model $M_{s\ell}[S_1][L]$. Therefore, the final model $M_{s\ell}[S_1][L][S_2]$ obtained from $M_{s\ell}[S_1][L]$ via the public announcement of S_2 is again unchanged:

$$M_{s\ell}[S_1][L][S_2] = M_{s\ell}[S_1][L].$$

As a result, Figure B11 depicts the final model obtained at the conclusion of the conversation. From this we obtain the solution: the numbers are 2 and 3.

3.3 Formalizing the Solution

Formalization of the Sum and Least Common Multiple Puzzle has certain similarities to the formalization of Cheryl's Birthday, though there are of course some differences. The formalization is easy once one has seen the formalization of the more difficult *Sum and Product Puzzle*. For details, we refer the reader to Plaza's (1989, 2007) formalization of the Sum and Product Puzzle.

Appendix C: Properties of binary relations

Below are the definitions of various adjectives that may be used to describe a binary relation R on a set W (of "worlds"): this is a set $R \subseteq W$ for which we write wRv to mean that $(w, v) \in R$.

- **Reflexive**: xRx for each $x \in W$.
  "Each world has an R-arrow pointing from that world right back to that world."
- **Transitive**: xRy and yRz implies xRz for each $x, y, z \in W$.
  "Every sequence of worlds connected by R-arrows gives rise to an R-arrow going directly from the first world to the last."
- **Symmetric**: xRy implies yRx for each $x, y \in W$.
  "Every R-arrow going in one direction gives rise to another R-arrow going in the opposite direction."
- **Equivalence relation**: R is reflexive, transitive, and symmetric.
- **Euclidean**: xRy and xRz implies yRz for each $x, y, z \in W$.
  "Two R-arrows leaving from the same source world give rise to R-arrows going between their destinations in both directions."
- **Serial**: for every $x \in W$, there is a $y \in W$ such that xRy.
  "Every world has a departing R-arrow."
- **Connected**: for each partition of W into two disjoint nonempty sets S and T, there is an $x \in S$ and a $y \in T$ such that xRy or yRx.
  "All worlds are linked to the same R-arrow network."
- **Complete (or Total)**: for any two different elements $x \in W$ and $y \in W$, we have xRy or yRx.
  "Every pair of distinct worlds is 'R-comparable'." (Think of xRy as saying, "x is less than y.")
- **Well-founded**: every nonempty set $S \subseteq W$ has an $x \in S$ such that for no $y \in S$ is it the case that yRx.
  "Every nonempty set of worlds has an 'R-minimal' element."
- **Converse well-founded**: the converse relation R^{-} (defined by $xR^{-}y$ iff yRx) is well-founded.
  "Every nonempty set of worlds has an 'R-maximal' element."

- **Preorder**: R is reflexive and transitive.

For information on the role of some of these properties in modal logic, we refer to the reader to our Appendix D or the *Stanford Encyclopedia of Philosophy* entry on Modal Logic.

Appendix D: Normal modal logic

Given a finite nonempty set $\mathcal{A}$ of "agents,' a *normal multi-modal logic with modals* $[a]$ *for each* $a \in \mathcal{A}$ is a theory that contains each of the following axiom schemes and rules:

- CL. Axiom schemes for classical propositional logic
- K. $[a](F \to G) \to ([a]F \to [a]G)$ for each $a \in \mathcal{A}$
- *Modus Ponens*: from $F \to G$ and F, infer G
- Modal Necessitation Rule: from F, infer $[a]F$ for any $a \in \mathcal{A}$

and may also include one or more of the following optional axiom schemes for one or more $a \in \mathcal{A}$:

- T. $[a]F \to F$
- D. $\neg[a]\bot$
- 4. $[a]F \to [a][a]F$
- 5. $\neg[a]F \to [a]\neg[a]F$

The basic normal multi-modal logic K is the minimal normal modal logic; in particular, it contains none of the optional rules. Each of the optional schemes corresponds to a specific property of the corresponding binary relation of a Kripke model; see Blackburn et al. (2002) for technical definitions and details. Scheme T corresponds to reflexivity of R_a, Scheme D corresponds to seriality of R_a, Scheme 4 corresponds to transitivity of R_a, and Scheme 5 corresponds to Euclideanness of R_a. The traditional "logic of knowledge" is $\mathsf{S5}$ (i.e., K plus optional schemes T, 4, and 5), while the traditional "logic of belief" is KD45 (i.e., K plus optional schemes D, 4, and 5). Other common normal modal logics include S4 (i.e., K plus optional schemes T and 4) and T (i.e., K plus optional scheme T). See Appendix C for definitions of the aforementioned relational properties.

For more on normal modal logics, we refer the reader to Blackburn et al. (2002) or the *Stanford of Encyclopedia Entry* entry on Modern Origins of Modal Logic.

Appendix E: Technical details of Public Announcement Logic

In this appendix, we take a closer look at some of the technical details of Public Announcement Logic.

- 1. Schematic validity
- 2. Expressivity and succinctness
- 3. Gerbrandy-Groeneveld announcements
- 4. Consistency-preserving announcements and Arrow Update Logic
- 5. Arbitrary Public Announcement Logic: quantification over announcements

---

1. Schematic validity

Fix a language $\mathcal{L}$ and a semantics for this language. To say that an $\mathcal{L}$-formula F is *schematically valid* means that F is valid and that F remains valid whenever we obtain a new $\mathcal{L}$-formula by replacing, for each propositional letter in F, all occurrences of that letter by some other $\mathcal{L}$-formula. To say that $\mathcal{L}$ itself is *schematically valid* means that every valid formula in $\mathcal{L}$ is schematically valid. (Validity may be restricted to some specific class of models; however, if no class is mentioned, then the class

of all models is assumed, as we will do now.) For example, let $\mathcal{L}$ be the language (ML) of basic multi-modal logic. Then the formula

$$p \to (q \to p)$$

is schematically valid: no matter what (ML)-formula we use to replace p or q, we always obtain a valid (ML)-formula. In fact, we can show that replacing letters in valid (ML)-formulas by (ML)-formulas preserves validity of *every* valid (ML)-formula; that is, (ML) is schematically valid. This proof is by induction on the length of derivations for multi-modal K: each axiom is schematically valid and each rule preserves schematic validity. Since an (ML)-formula is valid (for the class of all pointed Kripke models) if and only if it is derivable in multi-modal K, the result follows.

The language (PAL) is *not* schematically valid. In particular, the reduction axiom for propositional letters, which states that

$$[F!]p \leftrightarrow (F \to p) \text{ for letters } p \in \mathcal{P}$$

is not schematically valid. To see why, consider the formula obtained by replacing F with the letter q and p with the formula $[a]q$:

$$[q!][a]q \leftrightarrow (q \to [a]q) \text{ (not valid)}$$

This formula says that the statement "after the announcement of q, agent a knows q" is equivalent to the statement "if q is true, then agent a knows q". This supposed equivalence is formally and intuitively incorrect: an agent can come to know some truth after it is announced without knowing that truth in advance; that is, the left side $[q!][a]q$ can be true while the right side $q \to [a]q$ is false. Indeed, the primary purpose of a public announcement is to bring about knowledge by conveying information that the agents do not already have. The valid equivalence, which follows by Reduction Axiom 4, says something very different:

$$[q!][a]q \leftrightarrow (q \to [a][q!]q).$$

This says that a's knowledge of q after the announcement comes about if and only if, whenever q is true, a knows *before* the announcement that the possible future announcement of q would ensure that q is true. That this is correct follows immediately from the fact that PAL announcements are truthful and completely trustworthy: the announcement of a formula guarantees that the formula is true. This leads us to the key difference between the two equivalences: the valid one correctly equates post-announcement knowledge with pre-announcement knowledge of truths that a possible future announcement would convey, whereas the invalid one incorrectly equates post-announcement knowledge with what is currently known independent of what is to be announced.

Note that the axiomatization of PAL uses the letters F, G, and H to represent arbitrary formulas and the letter p to represent an arbitrary propositional letter. Formally speaking, each of these symbols is to be understood as a propositional letter, with the assumption that instances of a given axiom are obtained by replacing letters F, G, and H by (PAL)-formulas and the letter p by any letter. Therefore, to say that a given PAL axiom is schematically valid means that the axiom is valid (*i.e.,* each instance is valid) and, further, we obtain a valid formula whenever we replace one or more of F, G, H, or p by an *arbitrary* (PAL)-formula. In particular, this substitution need not respect the convention given in the axiomatization of PAL that restricts replacements of p to replacements by other letters. And herein is the source of non-schematic validity of (PAL): violating the requirement that p must be replaced by another letter allows us to replace p by a non-letter in the valid Reduction Axiom 1 and obtain a new, non-valid formula. Of course in making this replacement, the new

formula we obtain is not an instance of Reduction Axiom 1 (because we violated the replacement rule for p). Nevertheless, that we can obtain a non-valid formula from a valid one by letter replacement shows that (PAL) is not schematically valid.

Schematic validity is generally taken for granted when it comes to classical (modal) logic. It is therefore noteworthy that PAL, which is obtained as an extension of the language of classical multi-modal logic, is not schematically valid. In particular, one must take care when inferring that a formula F is valid in light of the fact that F has the same form as a known valid formula. We saw an example above where this can go wrong: from the valid Reduction Axiom 1 we obtained (by a replacement that violated this axiom's restriction on letters) a formula with the same form that turned out to be invalid. In general, writing $F(p/G)$ to denote the formula obtained from F by replacing all instances of p by G, the following is a rule that is not valid in PAL:

> **Schematic Substitution Rule (*not* valid for PAL):** from F, infer $F(p/G)$ for any formula G in the language.

But we do have the following kind of substitution rule in PAL:

> **Equivalence Substitution Rule (van Ditmarsch, van der Hoek, and Kooi 2007):** if $\mathsf{PAL} \vdash G \leftrightarrow H$, then $\mathsf{PAL} \vdash F(p/G) \leftrightarrow F(p/H)$.

Note that this rule is not part of the axiomatization of PAL; however, it can be proved that it is an admissible rule (i.e., the rule is derivable from the axiomatization of PAL, in the sense that if the hypotheses are PAL-derivable, then one can prove that the conclusion is as well).

Though not all (PAL)-formulas are schematically valid, there are some (PAL)-formulas that in fact are schematically valid. The formula $p \to (q \to p)$ is a simple example. An important result is that the set of schematically valid (PAL)-formulas—either over the class of transitive and Euclidean models with finitely many agents or over the class of all models with infinitely many agents—is decidable and axiomatizable; that is, there is an algorithm that can identify whether a given (PAL)-formula is schematically valid, and the collection of schematically valid (PAL)-formulas can be captured via an axiomatic system (Holliday et al. 2010, 2012). Whether this result holds for other classes of models is open as of the time of writing.

2. Expressivity and succinctness

The Reduction Theorem tells us that every (PAL)-formula F is provably equivalent to an announcement-free (ML)-formula F°. Put another way, every formula containing public announcements can be rewritten so as to express exactly the same thing without using any public announcements. Therefore, even though the language (PAL) of Public Announcement Logic is defined as an extension of the language (ML) of basic multi-modal logic, this extension does not let us say anything that could not already be said before.

> **PAL Expressivity Theorem (Plaza 1989, 2007; Gerbrandy and Groeneveld 1997).** (PAL) and (ML) are equally expressive (over the class of all pointed Kripke models). That is, whenever $\mathcal{L}$ is any one of our languages (PAL) and (ML) and $\mathcal{L}'$ is the other, we have the following: for each $\mathcal{L}$-formula F, there is an $\mathcal{L}'$-formula G such that $\vDash F \leftrightarrow G$.

This is a curious result: at first it seems like we have gained something by adding public announcements. In particular, the model-transforming operation of deleting worlds where the announced formula does not hold seems like something genuinely new; this is, after all, a *dynamic* take on what was before a purely *static* semantics. Nevertheless, the PAL

Expressivity Theorem tells us that, at least from the point of view of expressivity, new formulas for public announcements and an extended semantics to accommodate the corresponding model-transforming operation are not needed. This may seem like bad news for the PAL approach. But this is not the whole story, since using the announcement-free language to express public announcements comes with another cost.

> **PAL Succinctness Theorem (Lutz 2006; French et al. 2011, 2013).** (PAL) is exponentially more succinct than (ML) over the class $\mathcal{C}_{\mathsf{K}}$ of all Kripke models (Lutz 2006) and over the class $\mathcal{C}_{\mathsf{S5}}$ of reflexive, transitive, and Euclidean Kripke models (French et al. 2011, 2013): for each $\mathcal{C} \in \{\mathcal{C}_{\mathsf{K}}, \mathcal{C}_{\mathsf{S5}}\}$, there there is a sequence of (PAL)-formulas such that the size of the n-th formula F_n (in terms of the number of symbols) is a linear function of n but the smallest (ML)-formula F_n' satisfying $\mathcal{C} \vDash F \leftrightarrow F_n'$ has size at least 2^n.

This tells us that while we can express public announcement statements in the basic language (ML), doing so will in general require us to use exponentially more space. By way of analogy, we think of an advanced student of a foreign language: though she can express herself well in the foreign tongue, she may not have the ability to express every concept in the foreign language using a single word short phrase and so must sometimes use many words when speaking this language, even though she can make things much shorter in her mother tongue. So it is that the basic modal language (ML) can express fully the concepts of (PAL), though doing so will sometimes take exponentially more space.

This gives us two main reasons to stick with the PAL approach despite the seemingly bad news conveyed by the PAL Expressivity Theorem. First, using public announcements can save us an exponential amount of space. Second, the PAL approach provides us with a direct and intuitively appealing means of reasoning about truthful, completely trustworthy public announcements. That is, public announcements are *dynamic* operations that transform Kripke models (and hence agent knowledge and belief) in accordance with the information conveyed by what is announced, and using public announcements explicitly in the language can save us an exponential amount of space.

3. Gerbrandy-Groeneveld announcements

The model-transforming operation brought about by a public announcement of F was defined above as a *world-deleting* operation: delete all non-F worlds along with any arrows to or from these worlds. This method of *world-deleting* update was introduced by Plaza (1989, 2007). Gerbrandy and Groeneveld (1997) introduced an alternative *arrow-deleting* method: delete only the arrows pointing to non-F worlds. Defining new formulas $[F!_g]G$ with the intended meaning that G is true after the Gerbrandy–Groeneveld (GG) announcement of F, we are led to the following formal semantics:

- $M, w \models [F!_g]G$ holds if and only if $M[F!_g], w \models G$, where the model

$$M[F!_g] = (W[F!_g], R[F!_g], V[F!_g])$$

  is defined by:
  - $W[F!_g] := W$ — retain all worlds,
  - $xR[F!_g]_a y$ if and only if $xR_a y$ and $M, y \models F$ — delete arrows if and only if they point to non-F worlds, and
  - $v \in V[F!_g](p)$ if and only if $v \in V(p)$ — leave the valuation the same at all worlds.

Note that this semantics does not require that the announcement be truthful. That is, since we only delete arrows to non-F worlds, we do not run into a difficulty if we want to determine what is true of a non-F world w after the GG-announcement of F: we simply delete arrows to non-F

worlds (including those pointing to w) and then check what is true at w in the resulting model. This provides a natural way for us to drop the assumption made for Plaza-style world-deleting announcements that announcements are truthful. While this alternative may have some intuitive appeal, the two approaches turn out to be expressively equivalent: the language of modal logic (ML) with GG-announcements is just as expressive as the language (ML) itself, and hence it follows by the PAL Expressivity Theorem that (ML) with GG-announcements is just as expressive as (ML) with Plaza-announcements (= (PAL)).

In the remainder of the article, “public announcements” will refer to Plaza-announcements.

4. Consistency-preserving announcements and Arrow Update Logic

GG-announcements, like Plaza-announcements, maintain the assumption that announcements are completely trustworthy. For the Plaza approach, the agents’ trust of the announcement leads them to all together delete worlds that violate the announcement (along with arrows to and from these worlds). For the GG approach, the agents’ trust of the announcement leads them to all together eliminate arrows that point to a world that violates the announcement, effectively eliminating those worlds from the list of considered possibilities. In each case, a consequence of this complete trust is that the agents run into trouble when they receive inconsistent information. In particular, on announcement of a contradiction such as the propositional constant $\bot$ for falsehood, all agents’ beliefs become trivialized: the formulas $[\bot!][a]F$ and $[\bot!_g][a]F$ are valid for all agents a and all formulas F. In essence, an announcement of a contradiction brings each agent into a confused state of doxastic contradiction in which she believes anything and everything is true. This led Steiner (2006) to consider *consistency-preserving public* *announcements*, a variation of GG-announcements in which agents avoid arrow deletions that would lead to doxastic contradiction. Other studies of consistency preservation include Aucher (2008) and Aucher et al. (2009). Steiner's construction is a special case of a generalized approach to arrow deletion introduced by Kooi and Renne (2011a,b).

The language (AUL) of arrow update logic is given by the following grammar:

$$\begin{aligned} F &::= p \mid F \wedge F \mid \neg F \mid [a]F \mid [U]F \\ U &::= (F,a,F) \mid (F,a,F),U \\ & p \in \mathcal{P},\ a \in \mathcal{A} \end{aligned} \tag{AUL}$$

Each triple (G,a,G') is called an *a-arrow specification* with *source condition* G and *target condition* G'. A finite, comma-separated sequence U of arrow specifications for one or more agents is called an *arrow update*. When it is convenient, we identify an arrow update U with the finite set consisting of the arrow specifications in U.

An arrow update U describes the following Kripke model-transforming operation: save every a-arrow from a world x to a world x' if U contains an a-arrow specification (F,a,F') such that F is true at the source world x and F' is true at the target world x'; delete all other arrows in the model. Formally, this is defined as follows.

- $M,w \models [U]F$ holds if and only if $M * U, w \models F$, where the model

  $$M * U = (W * U, R * U, V * U)$$

  is defined by:
  - $W * U := W$ — retain all worlds,
  - $x(R * U)_a x'$ if and only if $xR_a x'$ and there exists $(G,a,G') \in U$ such that $M,x \models G$ and $M,x' \models G'$ — delete arrows that do not satisfy an arrow specification in U, and
  - $v \in V * U(p)$ if and only if $v \in V(p)$ — leave the valuation the same at all worlds.

Using the propositional constant $\top$ for truth, the GG-announcement formula $[F!_g]G$ is equivalent to the (AUL)-formula

$$[\{(\top,a,F) \mid a \in \mathcal{A}\}]G,$$

and the Plaza-announcement formula $[F!]G$ is equivalent to the (AUL)-formula

$$F \to [\{(\top,a,F) \mid a \in \mathcal{A}\}]G.$$

Arrow updates are therefore a generalization of both types of public announcement. Steiner's (2006) consistency-preserving public announcement with F is equivalent to Kooi and Renne's (2011a) "cautious update" with F, which is the arrow update CaU (called "CU" in Kooi and Renne 2011a) defined by

$$\mathsf{CaU}(F) := \{([a]\neg F,a,\top) \mid a \in \mathcal{A}\} \cup \{(\langle a\rangle F,a,F) \mid a \in \mathcal{A}\}.$$

Intuitively, $\mathsf{CaU}(F)$ says that if agent a believes that F is false, then a-arrows departing the current world should remain because she rejects an update she knows to be false. That is, we preserve a-arrows whose source is an $[a]\neg F$-world and whose target is a $\top$-world (i.e., whose target is any world at all). However, if she believes that F might be true, then all a-arrows leading to a non-F world should be deleted because she is willing to accept information that does not contradict her beliefs. That is, we preserve a-arrows whose source is an $\langle a\rangle F$-world only if their target is an F-world. We delete all arrows that do not satisfy one of the two kinds of arrow specifications that make up $\mathsf{CaU}(F)$.

A key limitation of arrow updates is that the information is assumed to be public: all agents together perform the same arrow update. So while arrow updates are a generalization of public announcements, this generalization still retains the assumption that the model-changing action is common knowledge among the agents. *Generalized Arrow Update Logic* (Kooi and Renne 2011b) drops this common knowledge assumption. We discuss Generalized Arrow Update Logic in Appendix I. Finally, we mention that Kooi and Renne (2011a) also present a *common knowledge* extension of (AUL), but this has not yet been extended to the generalized arrow update setting.

Having briefly surveyed consistency-preserving public announcements, Arrow Update Logic, and Generalized Arrow Update Logic, we return to the primary topic of this section: public announcements.

5. Arbitrary Public Announcement Logic: quantification over announcements

We may think of public announcement logic as a theory that tells us whether a given formula F is made true by some specific announcement G we have fixed in advance. But we might wish to look at this the other way around: is there some announcement G that will bring out the truth of our given formula F? Balbiani et al. (2008) introduced Arbitrary Public Announcement Logic (APAL) for studying such questions. The language (APAL) of Arbitrary Public Announcement Logic extends the language (PAL) of Public Announcement Logic by adding new “arbitrary announcement” formulas $[*!]F$ to express that “after every announcement, F is true”:

$$F ::= p \mid F \wedge F \mid \neg F \mid [a]F \mid [F!]F \mid [*!]F \qquad \text{(APAL)}$$
$$p \in \mathcal{P},\ a \in \mathcal{A}$$

The dual arbitrary announcement modality $\langle *!\rangle$ defined by setting

$$\langle *!\rangle F := \neg[*!]\neg F$$

allows us to express “after some announcement, F is true”. Semantically, the formula $[*!]F$ is interpreted as follows: to say that $[*!]F$ is true means that F is true after the announcement of any formula G of multi-modal logic (ML). Hence

$$[*!]F \text{ is intuitively equivalent to } \bigwedge_{G\in(\mathrm{ML})}[F!]F.$$

Of course the conjunction above is not a formula in our language (since it is infinite), but this expression gives the intuition of what we are after. In particular, the modal $[*!]$ acts as a *universal quantifier* over (ML)-formula announcements (“after every (ML)-announcement”), and the dual $\langle *!\rangle$ acts as an *existential quantifier* over (ML)-formula announcements (“after some (ML)-announcement”). Quantification is restricted to (ML)-formulas so as to avoid circularity in the semantics: were quantification to range over all (APAL)-formulas, $[*!]F$ would be true at a pointed model only if F holds after the formula $[*!]F$ is itself announced at that pointed model, and this would require us to already know the truth value of $[*!]F$ at that pointed model. Restricting the range of quantification to (ML)-formulas avoids this circularity. Further, we note it follows by the PAL Expressivity Theorem (Section 2) that restricting the range to (ML)-formulas is no different (in terms of expressivity) than restricting the range to (PAL)-formulas.

Formally, (APAL)-formulas are evaluated by extending the binary truth relation $\models$ between pointed Kripke models and (PAL)-formulas as follows:

- $M, w \models [*!]G$ holds if and only if we have $M, w \models [G!]F$ for all $G \in (\mathrm{ML})$.

Balbiani et al. (2008) obtained a number of results on the relative expressivity of (APAL), (PAL), and (ML).

Theorem (Balbiani et al. 2008). Over the class of all pointed $\mathsf{S5}$ Kripke models with agent set $\mathcal{A}$, we have each of the following.

- If $|\mathcal{A}| = 1$ (i.e., there is only one agent), then (APAL), (PAL), and (ML) are equally expressive.
- If $|\mathcal{A}| > 1$ (i.e., there is more than one agent), then (APAL) is strictly more expressive than both (PAL) and (ML).

The axiomatic theory of Arbitrary Public Announcement Logic is a deceptively simple extension of $\mathsf{S5}$ Public Announcement Logic obtained by adding one new axiom and one new rule. However, the proof of completeness is difficult and highly technical. We present the axiomatic theory, state the soundness and completeness results, and present two complexity results.

The axiomatic theory APAL.

- Axiom schemes and rules for PAL based on multi-agent $\mathsf{S5}$.
- $[*!]F \to [G!]F$ for $G \in$ (ML)
  "Arbitrary announcements may be instantiated by specific (ML)-formulas."
- Arbitrary Announcement Rule: from $F \to [G!][p!]H$ with p not occurring in F, G, or H, infer $F \to [G!][*!]H$.
  "If H obtains after an announcement by an otherwise unused letter p, then H obtains after an arbitrary announcement."

APAL **Soundness and Completeness (Balbiani et al. 2008).** APAL is sound and complete with respect to the collection C_* of pointed $\mathsf{S5}$ Kripke models. That is, for each (APAL)-formula F, we have that $\mathsf{APAL} \vdash F$ if and only if $C_* \models F$.

APAL Complexity. Let C_{S5} be the class of Kripke models such that each binary accessibility relation is reflexive, transitive, and symmetric.

- The satisfiability problem for (APAL) over C_{S5} is undecidable (French and van Ditmarsch 2008).
- The model checking problem for (APAL) over C_{S5} is PSPACE-complete (Balbiani et al. 2008).

APAL provides a convenient language for reasoning about Fitch's paradox (Fitch 1963), which concerns the seemingly strange result that the existence of unknown truths implies not all truths are knowable. Following the suggestion of van Benthem (2004), Balbiani et al. (2008) identify "F is knowable by agent a" with $\langle *!\rangle[a]F$. This equates "knowability" with "being known after some announcement". We then may express "all truths are knowable" using the scheme $F \to \langle *!\rangle[a]F$. We note that this scheme is not valid: the formula $p \wedge \neg[a]p$ ("p is true but a does not know it") is satisfiable but can never be known by a (because a's knowledge of the first conjunct p implies the negation of the second conjunct $\neg[a]p$). Therefore, if p is a truth unknown to agent a, then the formula $p \wedge \neg[a]p$ is another truth unknown to agent a but agent a cannot come to know this lattermost truth. That is, the following assumptions are jointly inconsistent:

- $p \wedge \neg[a]p$ — "p is an unknown truth."
- $F \to \langle *!\rangle[a]F$ for each $F \in$ (APAL) — "All truths are knowable."

We refer the reader to the discussions in van Benthem (2004), Balbiani et al. (2008), and Brogaard and Salerno (2012) for further details. Also, for an alternative interpretation of "all truths are knowable" based on the (APAL)-formula $\langle *!\rangle([a]F \vee [a]\neg F)$ ("it is possible for a to come to know *whether* F is true"), see van Ditmarsch et al. (2012b).

Ågotnes et al. (2010) study a variant of APAL known as “group announcement logic”. In this logic, new modal formulas having the form $\langle B!\rangle F$ for a group $B \subseteq \mathcal{A}$ of agents express that “the group B can make an announcement that makes F true”. The semantics of these new formulas is given as follows:

- $M, w \vDash \langle B!\rangle F$ holds if and only if we have $M, w \vDash \langle \bigwedge_{a\in B}[a]G_a!\rangle F$ for some function $G : B \to (\mathsf{ML})$ mapping each agent $a \in B$ to a formula $G_a \in (\mathsf{ML})$.

That is, to say that $\langle B!\rangle F$ is true means that there is a collection of (ML)-formulas G_a, one formula for each agent a in the group B, such that if each agent knows her formula and this knowledge is then announced, then F is true. For details, the reader is referred to Ågotnes et al. (2010).

Appendix F: The axiomatic theories of Relativized Common Knowledge

We base this theory, just as for PAL, on a PAL-friendly logic L. Here we let (L+RCK+P) be the extension of (L+PAL) obtained by adding relativized common knowledge formulas.

The axiomatic theory RCK+P.

- Axiom schemes and rules for a PAL-friendly logic L
- Reduction axioms (all in the language (L+RCK+P)):
  1. $[F!]p \leftrightarrow (F \to p)$ for letters $p \in \mathcal{P}$
     “After a false announcement, every letter holds—a contradiction. After a true announcement, letters retain their truth values.”
  2. $[F!](G \wedge H) \leftrightarrow ([F!]G \wedge [F!]H)$
     “A conjunction is true after an announcement iff each conjunct is.”
  3. $[F!]\neg G \leftrightarrow (F \to \neg[F!]G)$
     “G is false after an announcement iff the announcement, whenever truthful, does not make G true.”
  4. $[F!][a]G \leftrightarrow (F \to [a][F!]G)$
     “a knows G after an announcement iff the announcement, whenever truthful, is known by a to make G true.”
  5. $[F!][B*](G|H) \leftrightarrow (F \to [B*]([F!]G \mid F \wedge [F!]H))$
     “G is H-relative common knowledge among B after an announcement iff the announcement, whenever truthful, ensures that, relative to the announcement being true and H holding after the announcement, it is common knowledge among B that the announcement makes G true.”
- Announcement Necessitation Rule: from G, infer $[F!]G$ whenever the latter is in (L+RCK+P).
  “A validity holds after any announcement.”
- Axiom schemes for relativized common knowledge:
  - $[B*](F \to G \mid H) \to ([B*](F|H) \to [B*](G|H))$
    “Relativized common knowledge is closed under logical consequence.”
  - $[B*](F|H) \leftrightarrow [B](H \to (F \wedge [B*][F|H]))$, the “Mix axiom”
    “Relativized knowledge is equivalent to group knowledge that the relativization implies truth and relativized common knowledge.”
  - $([B](H \to F) \wedge [B*](F \to [B](H \to F) \mid H))$, the
    $\to [B*](F|H)$
    “Induction axiom”
    “If there is common knowledge that truth implies group knowledge and there is truth, then there is common knowledge.”
- RCK Necessitation Rule: from F, infer $[B*](F|H)$

"There is relativized common knowledge of every validity."

The axiomatic theory RCK is obtained from the axiomatic theory of RCK+P by omitting the reduction axioms and restricting formulas to the language (RCK).

RCK **Soundness and Completeness (van Benthem, van Eijck, and Kooi 2006).** RCK is sound and complete with respect to the collection $\mathcal{C}_*$ of pointed Kripke models for which the underlying modal logic L is sound and complete. That is, for each (RCK)-formula F, we have that RCK $\vdash F$ if and only if $\mathcal{C}_* \models F$.

As per the design, every formula in the language [RCK+P] of relativized common knowledge can be rewritten in an equivalent announcement-free form.

RCK+P Reduction Theorem (van Benthem, van Eijck, and Kooi 2006). Given a PAL-friendly theory L, every F in the language (L+RCK+P) of relativized common knowledge with public announcements is RCK+P-provably equivalent to a formula F° coming from the announcement-free language (RCK) of relativized common knowledge.

In the same way that the completeness of PAL was reduced to the completeness of the underlying modal logic L, the soundness of RCK+P and the Reduction Theorem for RCK+P together reduce the completeness of RCK+P to the completeness of the announcement-free underlying logic RCK.

RCK+P **Soundness and Completeness (van Benthem, van Eijck, and Kooi 2006).** Given a PAL-friendly theory L, the theory RCK+P is sound and complete with respect to the collection $\mathcal{C}_*$ of pointed Kripke models for which the underlying modal logic L is sound and

complete. That is, for each (L+RCK+P)-formula F, we have that RCK+P $\vdash F$ if and only if $\mathcal{C}_* \models F$.

The Reduction Theorem also tells us that adding public announcements to the language (RCK) of relativized common knowledge does not let us express anything that we could not already express before. We state this result formally in the following theorem along with a few other results that relate the languages we have looked at thus far in terms of expressivity.

Relative Expressivity Theorem. Over the class of all pointed Kripke models, we have each of the following.

- (RCK+P) and (RCK) are equally expressive (van Benthem, van Eijck, and Kooi 2006).
- (RCK) is strictly more expressive than (PAL+C): $[C](\neg[a]p \mid p)$ is equivalent to no (PAL+C)-formula (van Benthem, van Eijck, and Kooi 2006).
- (PAL+C) is strictly more expressive than (PAL): $[p!][C]q$ is equivalent to no (PAL)-formula (Baltag, Moss, and Solecki 1999; see also van Ditmarsch, van der Hoek, and Kooi 2007).
- (PAL) and (ML) are equally expressive (Plaza 1989, 2007; Gerbrandy and Groeneveld 1997).

A useful general reference on completeness via reduction axioms is Kooi (2007).

Appendix G: More on action models and the Logic of Epistemic Actions

In this Appendix, we examine three key topics in the Logic of Epistemic Actions: a notion of action model equivalence known as *action emulation*, how to extend action models so as to allow "factual change" via valuation-changing substitutions, and how to add common knowledge to the Logic of Epistemic Actions.

- 1. Action emulation
- 2. Incorporating factual change
- 3. Adding common knowledge

---

1. Action emulation

Two different pointed Kripke models can satisfy exactly the same (ML)-formulas, making them *semantically equivalent*. This lead to the study of a notion of *structural equivalence* for pointed Kripke models that would hold if and only if semantic equivalence holds as well, thereby allowing us to determine if two pointed Kripke models are semantically equivalent by examining only their structure. An often-used notion of structural equivalence is *bisimulation*

Bisimulation. A *bisimulation* between Kripke models $M_1 = (W_1, R_1, V_1)$ and $M_2 = (W_2, R_2, V_2)$ is a nonempty binary relation $Z \subseteq W_1 \times W_2$ between W_1 and W_2 satisfying the following properties:

- Propositional Identity: xZy implies $x \in V_1(p)$ if and only if $y \in V_2(p)$ for each letter $p \in \mathcal{P}$;
- Back: if xZy and $y(R_2)_a y'$, then there exists $x' \in W_1$ such that $x(R_1)_a x'$ and $x'Zy'$; and
- Forth: if xZy and $x(R_1)_a x'$, then there exists $y' \in W_2$ such that $y(R_2)_a y'$ and $x'Zy'$.

To say that the pointed Kripke models (M_1, w_1) and (M_2, w_2) are *bisimilar*, written $(M_1, w_1) \leftrightarroweq (M_2, w_2)$, means there exists a bisimulation Z between M_1 and M_2 satisfying $w_1 Z w_2$. To say M_1 and

M_2 are *bisimilar*, written $M_1 \leftrightarroweq M_2$, means there exists a bisimulation Z between M_1 and M_2 that is *total*: every $x \in W_1$ has a $y \in W_2$ such that xZy, and every $y \in W_2$ has an $x \in W_1$ such that xZy.

Bisimulation Theorem (see, e.g., Blackburn, de Rijke, and Venema 2002). Let (M_1, w_1) and (M_2, w_2) be pointed Kripke models. We write $(M_1, w_1) \equiv (M_2, w_2)$ to indicate that these pointed Kripke models are (ML)-*semantically equivalent*, meaning we have $M_1, w_1 \models F$ if and only if $M_2, w_2 \models F$ for each (ML)-formula F.

- Bisimulation implies (ML)-semantic equivalence:

$$(M_1, w_1) \leftrightarroweq (M_2, w_2) \text{ implies } (M_1, w_1) \equiv (M_2, w_2).$$

- (ML)-semantic equivalence of image finite models implies bisimulation: if M_1 and M_2 are *image finite*, meaning each world in the model has at most finitely many outgoing R_a-arrows for each agent $a \in \mathcal{A}$, then

$$(M_1, w_1) \equiv (M_2, w_2) \text{ implies } (M_1, w_1) \leftrightarroweq (M_2, w_2).$$

Notice that the Bisimulation Theorem does not give us an exact correspondence between bisimulation and (ML)-semantic equivalence: the second item of the theorem requires the models be image finite. A more complicated version of the Bisimulation Theorem makes it possible to avoid the issue of image finiteness; see Blackburn et al. (2002) for details.

Similar to the way in which different pointed Kripke models can be (ML)-semantically equivalent, different actions (A_1, e_1) and (A_2, e_2) can also be "equivalent". Following van Eijck, Ruan, and Sadzik (2012), action equivalence is defined in terms of bisimulation: to say that two actions are "equivalent" means that the actions can be executed in the same situations and, furthermore, whenever the actions are both executable, they yield bisimilar (and hence (ML)-indistinguishable) results.

> **Equivalence of action models (van Eijck, Ruan, and Sadzik 2012).** To say that action models A_1 and A_2 are *equivalent*, written $A_1 \equiv A_2$, means that
>
> - A_1 and A_2 are executable in the same Kripke models; and
> - if A_1 and A_2 are both executable in the Kripke model M, then $M[A_1] \leftrightarroweq M[A_2]$.
>
> To say that the actions (A_1, e_1) and (A_2, e_2) are *equivalent*, written $(A_1, e_1) \equiv (A_2, e_2)$, means that the underlying action models A_1 and A_2 are equivalent (i.e., $A_1 \equiv A_2$).

Figure G1 provides an example of equivalent actions. Examining this figure, it is clear that A_1 and A_2 are executable in the same situations: each is executable in situations satisfying p and in situations satisfying $\neg p$. Further, these two action models produce bisimilar results: events e_1 and f_1 execute in the same situations and events f_2 and f_3 effectively "simulate" the always-executable event e_2. This "simulation" works as follows: any situation at which e_2 is executable satisfies either p or $\neg p$ and so may be matched up with one of f_2 or f_3. Since all events are interconnected in both action models, the connections between e_2-produced worlds match up with the connections between the corresponding f_2- or f_3-produced worlds in a way that ensures bisimilar results. For example, for a sequence of three worlds w, x, y, z connected in order using R_a-arrows and respectively satisfying p, p, $\neg p$, and p in the original model:

- the A_1-produced (w, e_1), (x, e_2), (y, e_2), and (z, e_1) are connected in order using R_a-arrows; and
- the A_2-produced final model will have the worlds (w, f_1), (x, f_2), (y, f_3), and (z, f_2) are connected in order using R_a-arrows.

Sietsma and van Eijck (2012) provide a full proof that these actions are equivalent.

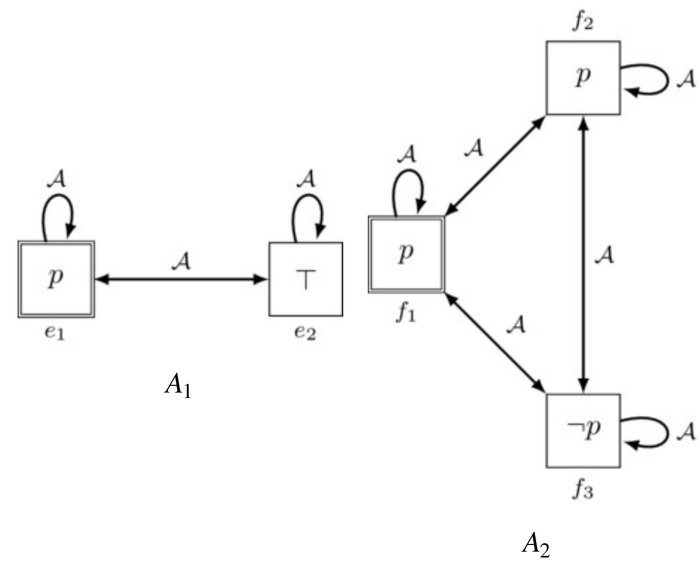


Figure G1. The actions (A_1, e_1) and (A_2, f_2) are equivalent (van Eijck, Ruan, and Sadzik 2012; Sietsma and van Eijck 2012).

It is possible to define a notion of bisimulation for actions, but it turns out that there are equivalent actions that are not bisimilar (van Eijck, Ruan, and Sadzik 2012; Sietsma and van Eijck 2012). This led to the search for a structural relationship between actions that hold if and only if the actions are equivalent. A notion of *parameterized action emulation* was proposed by van Eijck, Ruan, and Sadzik (2012), and a variant called *action emulation* was later proposed by Sietsma and van Eijck (2012). To introduce these, we first need the following preliminary definitions.

> **Single negation, closure, atoms, and $\mathsf{pre}(A)$ (Sietsma and van**

Eijck 2012). Let S be a set of (ML)-formulas.

- The *single negation* of an (ML)-formula F, written $\sim F$, is defined as follows: if F has the form $\neg G$, then $\sim F$ is G; otherwise, $\sim F$ is $\neg F$. The *single negation* of the set S, written $\sim S$, is the set consisting of the single negation of every formula in S.
- The *closure* of S is the smallest set $\mathsf{Cl}(S)$ that contains S and is closed under subformulas (i.e., $F \in \mathsf{Cl}(S)$ and G a subformula of F together imply $G \in \mathsf{Cl}(S)$) and closed under single negations (i.e., $\sim\mathsf{Cl}(S) = \mathsf{Cl}(S)$).
- An *atom* over S is a maximal consistent subset of $\mathsf{Cl}(S)$: a subset T of $\mathsf{Cl}(S)$ that is *consistent* (i.e., $\not\models \neg \bigwedge T$) and that satisfies the property that adding any formula in $\mathsf{Cl}(S)$ that is not already present will result in a set that is *inconsistent* (i.e., not consistent). We let $\mathsf{At}(S)$ denote the set of atoms over S.
- For each action model A, define the set $\mathsf{pre}(A) := \{\mathsf{pre}(e) \mid e \in E^A\}$ of preconditions in A.
- A *multi-pointed action model* is a pair (A, Z) consisting of an action model $A = (E, A, \mathsf{pre})$. and a set $Z \subseteq E$ of events in A called the *multi-point*. We let

  $$[A, Z]F \quad \text{abbreviate} \quad \textstyle\bigwedge_{e \in Z}[A, e]F,$$

  so that a multi-pointed action model modality acts as a conjunction over all events in the multi-point. Note that the conjunction is equal to $\top$ if the multi-point Z is empty. Multi-pointed action models will sometimes be called *multi-actions*. If we write a pointed action model (A, e) where a multi-pointed action model is expected, then what is meant is the multi-action $(A, \{e\})$. In this way, actions may be identified with multi-actions.

The single negation of a formula strips any outermost negation that is present or, if none is present, adds one. The single negation of a set just performs the single negation of each formula in the set. The closure of a set expands the set by repeatedly adding subformulas and performing single negations on the formulas present until no further expansion is possible. In essence, the closure if a set of formulas is the collection of all formulas that are “relevant” when it comes to determining whether a formula in the set is true at a pointed Kripke model. An atom over a set of formulas is a subset whose formulas can all be made true at a pointed Kripke model and, further, any larger subset would violate this property (see, e.g., Blackburn, de Rijke, and Venema 2002). In effect, choosing an atom of a set S is the same as choosing a pointed Kripke model that makes true the formulas in the atom and makes false the other formulas in S that are not in the atom. By choosing the set $S = \mathsf{pre}(A_1) \cup \mathsf{pre}(A_2)$ to be the collection of all precondition formulas in either the action model A_1 or the action model A_2, an atom $\sigma \in \mathsf{At}(S)$ over S identifies a maximal collection of preconditions in these action models that can simultaneously be made true: these are just the actions e having $\mathsf{pre}(e) \in \sigma$. These ideas lead to the following notion of structural equivalence between action models.

> **Parameterized action emulation (adapted from van Eijck, Ruan, and Sadzik 2012).** Let $A_1 = (E_1, R_1, \mathsf{pre}_1)$ and $A_2 = (E_2, R_2, \mathsf{pre}_2)$ be action models and let $S = \mathsf{pre}(A_1) \cup \mathsf{pre}(A_2)$ be the set of preconditions for all events in either of A_1 or A_2. For each $i \in \{1, 2\}$ and event $e \in E_i$, define
>
> $$\Sigma_i(e) := \{\sigma \in \mathsf{At}(S) \mid \mathsf{pre}_i(e) \in \sigma\}$$
>
> to be the set of atoms satisfying the precondition of e. A *parameterized action emulation* between A_1 and A_2 is a set of indexed binary relations $\{E_\sigma\}_{\sigma \in \mathsf{At}(S)}$ between E_1 and E_2 (i.e., $E_\sigma \subseteq E_1 \times E_2$

for each $\sigma \in \mathsf{At}(S)$) satisfying the property that $xE_\sigma y$ implies the following:

- Invariance: $\mathsf{pre}(x) \in \sigma$ and $\mathsf{pre}(y) \in \sigma$;
- Back: if we have that $y(R_2)_a y'$, that $\sigma \in \Sigma_2(y')$ and that $\bigwedge \sigma \wedge \langle a \rangle \bigwedge \sigma'$ is consistent, then there exists $x' \in E_1$ such that $x(R_1)_a x'$ and $x' E_{\sigma'} y'$; and
- Forth: if we have that $x(R_1)_a x'$, that $\sigma \in \Sigma_2(x')$ and that $\bigwedge \sigma \wedge \langle a \rangle \bigwedge \sigma'$ is consistent, then there exists $y' \in E_2$ such that $y(R_1)_a y'$ and $x' E_{\sigma'} y'$.

For multi-actions (A_1, Z_1) and (A_2, Z_2), we write $(A_1, Z_1) \rightleftarrows_p (A_2, Z_2)$ to mean that there is a parameterized action emulation $\{E_\sigma\}_{\sigma \in \mathsf{At}(S)}$ between A_1 and A_2 satisfying the following:

- for each $x \in Z_1$ and $\sigma \in \Sigma_1(x)$, there exists $y \in Z_2$ such that $xE_\sigma y$; and
- for each $y \in Z_2$ and $\sigma \in \Sigma_2(x)$, there exists $x \in Z_1$ such that $xE_\sigma y$.

Intuitively, an emulation between actions links events e and f via $eE_\sigma f$ only if the preconditions of these events can be made simultaneously true in a situation that makes true all formulas in the atom σ. Further, this linkage his bisimulation-like Back and Forth properties that respect consistency of the linking atoms: a link made between events via E_σ gives rise to another link made between events via E_σ only if it is consistent for a situation satisfying the formulas in σ to lead via an a-arrow to a situation satisfying the formulas in σ' (i.e., $\bigwedge \sigma \wedge \langle a \rangle \bigwedge \sigma'$ is consistent). In this way, the Back and Forth conditions come with built-in guarantees of consistency, and this is leveraged in proving the following result.

Parameterized Action Emulation Theorem (van Eijck, Ruan, and Sadzik 2012). For multi-actions (A_1, Z_1) and (A_2, Z_2), we have

$$(A_1, Z_1) \leftrightarrows_p (A_2, Z_2) \text{ if and only if } A_1 \equiv A_2.$$

This gives a direct correspondence between a parameterized structural relationship over actions and the semantic notion of action equivalence. However, the parameterization makes this notion complex. A simpler, non-parameterized structural relationship that gets close to providing a similar direct correspondence was recently proposed by Sietsma and van Eijck (2012). This relationship and the related results are formulated in terms of *multi-pointed action models*.

Action emulation (adapted from Sietsma and van Eijck 2012). Let $A_1 = (E_1, R_1, \mathsf{pre}_1)$ and $A_2 = (E_2, R_2, \mathsf{pre}_2)$ be action models. An *action emulation* between A_1 and A_2 is a binary relation $E \subseteq E_1 \times E_2$ satisfying the property that xEy implies:

- Consistency: $\mathsf{pre}_1(x) \wedge \mathsf{pre}_2(y)$ is consistent;
- Back: $y(R_2)_a y'$ implies there exists $X' \subseteq E_1$ such that $x(R_1)_a x'$ and $x' E_a y'$ for each $x' \in X'$ and
$$\vDash (\mathsf{pre}_1(x) \wedge \mathsf{pre}_2(y)) \to [a](\mathsf{pre}_2(y') \to \bigvee_{x' \in X'} \mathsf{pre}_1(x'));$$
- Forth: $x(R_1)_a x'$ implies there exists $Y' \subseteq E_2$ such that $y(R_1)_a y'$ and $x' E_a y'$ for each $y' \in Y'$ and
$$\vDash (\mathsf{pre}_1(x) \wedge \mathsf{pre}_2(y)) \to [a](\mathsf{pre}_2(x') \to \bigvee_{y' \in Y'} \mathsf{pre}_1(y')).$$

For multi-actions (A_1, Z_1) and (A_2, Z_2), we write $(A_1, Z_1) \leftrightarrows (A_2, Z_2)$ to mean that there is an action emulation E between A_1 and A_2 satisfying the following:

- every $x \in Z_1$ gives rise to a $Y \subseteq Z_2$ such that xEy for each $y \in Y$ and

$$\vDash \mathsf{pre}_1(x) \to \bigvee_{y \in Y} \mathsf{pre}_2(y);$$

- every $y \in Z_2$ gives rise to a $X \subseteq Z_1$ such that xEy for each $x \in X$ and

$$\vDash \mathsf{pre}_1(y) \to \bigvee_{x \in X} \mathsf{pre}_2(X).$$

An action emulation connects events only if their preconditions are consistent. Further, for events e and f linked via eEf, an event reachable from one of e and f via a-arrows is linked up with a set of events reachable from the other via a-arrows in such a way that the preconditions of e and f together imply that a-accessible worlds must satisfy a precondition from an event in the set whenever the precondition from the a-accessible event is satisfied. As it turns out, this is just what is needed to show that emulation implies equivalence. The converse (i.e., the statement that equivalence implies emulation) is an open question. However, a weaker version of the converse has been proved. This version makes use of the following result.

Canonical Action Model Theorem (Sietsma and van Eijck 2012). Let S be a finite set of (ML)-formulas and $A = (E, R, \mathsf{pre})$ be an action model. To say that A is *canonical* over S means that:

- for each $e \in E$, there is an atom $\sigma \in \mathsf{At}(S)$ over S such that $\mathsf{pre}(e) = \bigwedge \sigma$, and
- $eR_a e'$ implies $\mathsf{pre}(e) \wedge \langle a \rangle \mathsf{pre}(e')$ is consistent.

To say that an action or multi-action is *canonical* means that the action model underlying (i.e., the first element of the pair making up) the action or multi-action is canonical. The multi-action $(A, \{e\})$ is equivalent to the multi-action (A^c, e^c), where the canonical action model $A^c := (E^c, R^c, \mathsf{pre}^c)$ over $\mathsf{pre}(A)$ and the event $e^c \in E^c$ are defined as follows:

- $E^c := \{(x, \sigma) \in E \times \mathsf{At}(\mathsf{pre}(A)) \mid \mathsf{pre}(x) \in \sigma\}$,
- $(x, \sigma) R^c_a (x', \sigma')$ if and only if $xR_a x'$ and $\bigwedge \sigma \wedge \langle a \rangle \bigwedge \sigma'$ is consistent,
- $\mathsf{pre}^c((x, \sigma)) := \bigwedge \sigma$, and
- $e^c := \{(x, \sigma \in E^c \mid x = e\}$.

An action model that is canonical over a set S of formulas is an action model whose preconditions are conjunctions of atoms over S. Therefore, each precondition describes a situation in which the formulas in the corresponding atom over S are all true and the other formulas in S are all false. This gives important information that is exploited in proof of the following result.

Action Emulation Theorem (Sietsma and van Eijck 2012). Let (A_1, Z_1) and (A_2, Z_2) be multi-actions.

- Emulation implies equivalence:

$$(A_1, Z_1) \leftrightarrows (A_2, Z_2) \quad \text{implies} \quad A_1 \equiv A_2;$$

- Equivalence between canonical action models implies emulation of multi-actions: if A_1 is canonical over S_1 and A_2 is canonical over S_2, then

$$A_1 \equiv A_2 \quad \text{implies} \quad (A_1, Z_1) \leftrightarrows (A_2, Z_2).$$

Non-parameterized action emulation provides a procedure for determining whether two actions (A_1, e_1) and (A_2, e_2) are equivalent: construct the equivalent canonical multi-actions (A^c_1, e^c_1) and (A^c_2, e^c_2) and determine whether

$$(A^c_1, e^c_1) \leftrightarrows (A^c_2, e^c_a).$$

Since A_1^c and A_2^c are both finite and computable from A_1 and A_2, it follows that the overall procedure is computable as well.

It is an open question whether the canonicity condition can be dropped in the second part of the Action Emulation Theorem. If this condition cannot be dropped, then it is possible that there may be a variant of the above-defined notion of non-parameterized action emulation that would correspond directly with action equivalence (as in the Parameterized Action Emulation Theorem except without the parameterization).

2. Incorporating factual change

As a consequence of the (EAL) semantics presented in the main article, an executable action (A,e) transforms an initial world w into a resultant world (w,e) in a manner that preserves the valuation. That is, a propositional letter p is true at the resultant world (w,e) if and only if p was true at the initial world w. By the soundness of the axiomatic system EAL, the preservation of propositional valuation must be reflected in the axiomatics as well. And indeed, this is so: the EAL reduction axiom for propositional letters

$$[A,e]p \leftrightarrow (\mathsf{pre}(e) \to p) \text{ for letters } p \in \mathcal{P}$$

implies the scheme

$$\mathsf{pre}(e) \to ([A,e]p \leftrightarrow p) \text{ for letters } p \in \mathcal{P},$$

which says that whenever (A,e) is executable, p is true after the occurrence of (A,e) if and only if p was true before the occurrence. Since this is true of all propositional letters, it follows that all actions are valuation-preserving. Thinking of propositional letters as the "basic facts" of the situations that we reason about using EAL, the preservation of valuations is sometimes described as "preserving (basic) facts" or as "leaving (basic) facts unchanged".

Sometimes we want to reason about fact-changing actions. For example, if we use the letter q to represent the statement that "the light is on" and this is true in an initial world w, then the action of "flipping the light switch" should bring about a resultant world (w,e) in which q is false. That is, the valuation should change: the basic fact q goes from being true in w to being false in (w,e).

Incorporating factual change into the theory of action models requires two modifications: one syntactical-semantical and one axiomatic. For the syntactical-semantical modification, we add an extra component to action models that allows us to specify changes in the valuation of various letters.

> **Action model with substitution (van Benthem, van Eijck, and Kooi 2006).** Given a set of formulas $\mathcal{L}$ and a finite nonempty set $\mathcal{A}$ of agents, an *action model with substitution* is a structure $A = (E, R, \mathsf{pre}, \mathsf{sub})$ satisfying the following:
>
> - (E, R, pre) is an action model (as defined before); and
> - $\mathsf{sub} : (E \times \mathcal{P}) \to \mathcal{L}$ is a *(finite) substitution*: a function mapping each event-letter pair $(e,p) \in E \times \mathcal{P}$ to an $\mathcal{L}$-formula $\mathsf{sub}(e,p)$ subject to the restriction that $\mathsf{sub}(e,p) = p$ for all but finitely many pairs $(e,p) \in E \times \mathcal{P}$.

We then define the language (EAL+S) of Epistemic Action Logic with substitution along with the set $\mathsf{AM{+}S}_*$ of pointed action models with substitution having preconditions in the language (EAL+S) much as we did for (EAL) and AM_* but with a slightly different recursive grammar (in which pointed action models come from $\mathsf{AM{+}S}_*$ instead of from AM_*):

$$F ::= p \mid F \wedge F \mid \neg F \mid [a]F \mid [A,e]F \qquad \text{(EAL+S)}$$
$$p \in \mathcal{P},\ a \in \mathcal{A},\ (A,e) \in \mathsf{AM{+}S}_*$$

A substitution sub specifies substitutions for propositional letters: to have $\mathsf{sub}(e,p) = F$ means that we are to reassign the truth value of p so that it is like F as per the occurrence of event e. That is, at any world w at which e can occur (i.e., at which the precondition $\mathsf{pre}(e)$ holds), the execution of e yields a world (w,e) at which p is true if and only if F was true before in w. As a result, p is essentially replaced with F during the occurrence of the event. Formally, this works as follows:

- $M,w \vDash [A,e]G$ holds if and only if $M,w \nvDash \mathsf{pre}(e)$ or $M[A],(w,e) \vDash G$, where the model

  $$M[A] = (W[A], R[A], V[A])$$

  is the result of the *product update (with substitution)* $M \mapsto M[A]$ defined by:
  - $W[A] := \{(v,g) \in W \times E \mid M,v \vDash \mathsf{pre}(g)\}$ — pair worlds with events whose preconditions they satisfy,
  - $(w_1,e_1)R[A]_a(w_2,e_2)$ if and only if there is an a-arrow from w_1 to w_2 in M and an a-arrow from e_1 to e_2 in A, and
  - $(v,g) \in V[A](p)$ if and only if $M,v \vDash \mathsf{sub}(g,p)$ — make the truth of p at (v,g) like that of the formula $\mathsf{sub}(g,p)$ at v.

So if $\mathsf{sub}(e,p) = F$, then the letter p is true after the occurrence of e if and only if F was true before the occurrence of e. Therefore, $\mathsf{sub}(e,p) = p$ implies that the valuation of p is left unchanged, which is how the semantics worked before we introduced substitutions. Our assumption that we have $\mathsf{sub}(e,p) = p$ for all but finitely many pairs (e,p) ensures that the valuation changes for at most finitely many letters. The reason for this restriction is that we want action models to be finite objects: it should be

possible in principle to write down a sequence of symbols (or draw a picture) that completely specifies any given action model in our language.

The axiomatic theory $\mathsf{EAL{+}S}$ of Epistemic Action Logic with substitutions is identical to the theory EAL except that the reduction axiom for propositional letters is changed to the following:

1. $[A,e]p \leftrightarrow (\mathsf{pre}(e) \to \mathsf{sub}(e,p))$ for letters $p \in \mathcal{P}$
   "After a non-executable action, every letter holds—a contradiction. After an executable action, the substitution sets truth values."

An example utilizing an action model with substitution is given in Figure G2, where the action $\mathsf{Flip}(q)$ implements our light-switch action: the occurrence of this action toggles the truth value of q between "on" (i.e., true) and "off" (i.e., false) in a manner that is common knowledge to all.

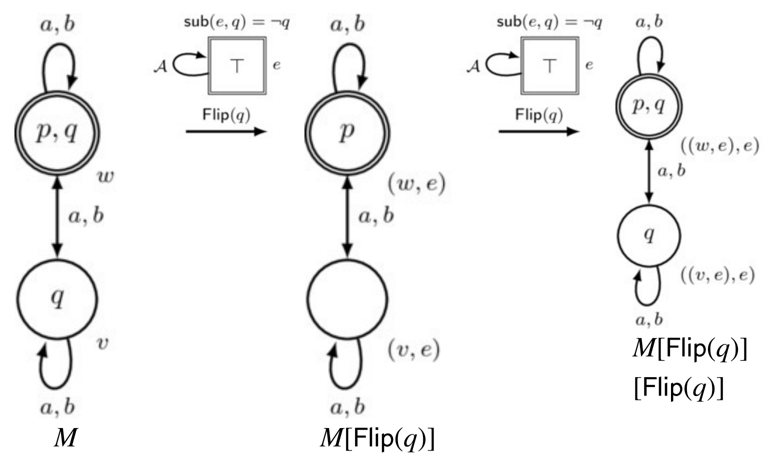


FIGURE G2. The pointed action model $(\mathsf{Flip}(q), e)$ toggles the truth value of letter q in a way that is common knowledge to all.

Almost all results about action model logics carry through *mutatis mutandis* upon addition of valuation-changing substitutions. So in the interest of simplicity, we assume in what follows that, unless stated otherwise, our action models do not contain valuation-changing substitutions. It is worth noting that this assumption is often made in the DEL literature and, in light of how simple it is to include valuation-changing substitutions if they are desired, this assumption should not be viewed as a genuine limitation to a given piece of work.

3. Adding common knowledge

To study the effect of common knowledge on action models, we define the language (EAL+C) of (EAL) with common knowledge along with the set $\mathsf{AM}{+}C_*$ of pointed action models with preconditions in the language (EAL+C) according to the following recursive grammar:

$$F ::= p \mid F \wedge F \mid \neg F \mid [a]F \mid [A,e]F \mid [B*]F \qquad \text{(EAL+C)}$$
$$p \in \mathcal{P},\ a \in \mathcal{A},\ (A,e) \in \mathsf{AM}{+}C_*,\ B \subseteq \mathcal{A}$$

The semantics of this language over pointed Kripke models is defined above. There are a number of interesting fragments of (EAL+C) obtained by restricting the action model modalities $[A,e]$ to have certain communicative forms. Here are a few examples.

- (PUB+C) is the fragment of (EAL+C) obtained by requiring that any action model modality $[A,e]$ used in the formation of a formula be a public announcement modality $[\mathsf{Pub}(F),e]$. See Figure 4 (main article) for a picture of the action model $\mathsf{Pub}(p)$.
- (PRI+C) is the fragment of (EAL+C) obtained by requiring that any action model modality $[A,e]$ used in the formation of a formula be a private announcement modality $[\mathsf{Pri}(F),e]$ Figure 3 (main article) depicts the action model $\mathsf{Pri}(p)$.

We now examine the axiomatic theory of EAL with common knowledge.

The axiomatic theory $\mathsf{EAL}{+}C$. Other name in the literature: AMC (van Ditmarsch, van der Hoek, and Kooi 2007).

- Axiom schemes and rules for the theory EAL
- Axiom schemes for common knowledge:
  - $[B*](F \to G) \to ([B*]F \to [B*]G)$
    "Common knowledge is closed under logical consequence."
  - $[B*]F \leftrightarrow (F \wedge [B][B*]F)$, the "Mix axiom"
    "Common knowledge is equivalent to truth and group knowledge of common knowledge."
  - $[B*](F \to [B]F) \to (F \to [B*]F)$, the "Induction axiom"
    "If there is common knowledge that truth implies group knowledge and there is truth, then there is common knowledge."
- CK Necessitation Rule: from F, infer $[B*]F$
  "There is common knowledge of every validity."
- AM-CK Rule: given (EAL+C)-formulas G_x for each event x in the action model A, from

  $$\bigwedge_{e(R_B)*f}\left((G_f \wedge \mathsf{pre}(f) \wedge (G_f \to [A,f]F)) \to \bigwedge_{a\in B}\bigwedge_{fR_a g}[a]G_g\right)$$

  infer

  $$G_e \to [A,e][B*]F$$

Just like for $\mathsf{PAL}{+}C$, there can be no reduction theorem for $\mathsf{EAL}{+}C$. This makes for a more intricate completeness proof.

$\mathsf{EAL}{+}C$ Soundness and Completeness (Baltag, Moss, and Solecki 1999; van Ditmarsch, van der Hoek, and Kooi 2007). $\mathsf{EAL}{+}C$ is sound and complete with respect to the collection C_* of pointed Kripke models for which the underlying logic EAL is sound and complete. That is, for each (EAL+C)-formula F, we have $\mathsf{EAL}{+}C \vdash F$ if and only if $C_* \vDash F$.

We summarize some of the known expressive relationships between various fragments of (EAL+C) over the class of all pointed Kripke models. Relationships over various other classes are reported by van Ditmarsch, van der Hoek, and Kooi (2007).

Relative Expressivity Theorem. Over the class of all pointed Kripke models, we have each of the following.

- (EAL+C) is strictly more expressive than each of (PUB+C) and (PRI+C) (Baltag, Moss, and Solecki 1999).
- (PUB+C) and (PRI+C) are expressively incomparable: there is a (PUB+C)-formula that cannot be expressed in (PRI+C) (Renne 2008), and there is a (PRI+C)-formula that cannot be expressed in (PUB+C) (Baltag, Moss, and Solecki 1999).
- (PUB+C) and (PAL+C) are equally expressive (Baltag, Moss, and Solecki 1999).
- (PAL+C) is strictly more expressive than (ML+C) (Baltag, Moss, and Solecki 1999).
- (ML+C) is strictly more expressive than (ML) (Fagin et al. 1995).
- (EAL) and (ML) are equally expressive (Baltag, Moss, and Solecki 1999).

In closing, we note two alternative approaches to reasoning about action models and common knowledge.

- van Benthem, van Eijck, and Kooi (2006) extend (ML) to an "Epistemic Propositional Dynamic Logic" that can express common knowledge (and relativized common knowledge). It is shown via a Reduction Theorem that adding action model modalities does not increase expressivity of the language.

- Van Benthem and Ikegami (2008) show that modal μ-calculus, which can express common knowledge, is closed under the product update. That is, a Reduction Theorem implies that adding action model modalities does not increase expressivity of the language.

Appendix H: Recursive definition of languages with action models

Formally, the grammar (EAL) is defined by double recursion as follows. First, let (EAL^0) be the language (ML) of modal logic, and let AM_*^0 be the set of pointed action models whose precondition qformulas all come from the language (EAL^0). Then, whenever (EAL^n) and AM_*^n are both defined, we define the language (EAL^{n+1}) and the set AM_*^{n+1} as follows:

$$F ::= p \mid F \wedge F \mid \neg F \mid [a]F \mid [A,e]F \qquad (\mathrm{EAL}^{n+1})$$
$$p \in \mathcal{P},\ a \in \mathcal{A},\ (A,e) \in \mathsf{AM}_*^n$$

and we let AM_*^{n+1} be the set of pointed action models whose precondition formulas all come from the language (EAL^{n+1}). Finally, we define (EAL) as the union

$$(\mathrm{EAL}) := \bigcup_{n\in\mathbb{N}} (\mathrm{EAL}^n)$$

of all languages (EAL^n) and we define AM_* to be the union

$$\mathsf{AM}_* := \bigcup_{n\in\mathbb{N}} \mathsf{AM}_*^n$$

of all sets AM_*^n. For convenience, we let AM^n denote the set of all action models whose precondition formulas are all in the language (EAL^n), and we let AM denote the set of all action models whose precondition formulas are all in the language (EAL). A useful lemma is the following.

Inclusion Lemma. $m \leq n$ implies $(\text{EAL}^m) \subseteq (\text{EAL}^n)$ and $\mathsf{AM}^m \subseteq \mathsf{AM}^n$.

We often need a complete, well-founded binary relation on (EAL)-formulas that orders these formulas in such a way that the right side of each of the following equivalences—the EAL reduction axioms—has a "smaller" position in the ordering than does the left side (so that going left to right in a reduction axiom "reduces" the position in the ordering).

1. $[A, e]p \leftrightarrow (\mathsf{pre}(e) \to p)$ for letters $p \in \mathcal{P}$
2. $[A, e](G \wedge H) \leftrightarrow ([A, e]G \wedge [A, e]H)$
3. $[A, e]\neg G \leftrightarrow (\mathsf{pre}(e) \to \neg[A, e]G)$
4. $[A, e][a]G \leftrightarrow (\mathsf{pre}(e) \to \bigwedge_{eR_a f}[a][A, f]G)$

A natural strict dictionary ordering is given as follows. First, for each [EAL]-formula F, let $d(F)$ denote the *(non–precondition-recursing) formation depth* of F and $D(F)$ denote the *post-action (non–precondition-recursing) formation depth* of F.

$$
\begin{aligned}
d(p) &= 0 \text{ for letters } p \in \mathcal{P} & D(p) &= 0 \text{ for letters } p \in \mathcal{P} \\
d(F \wedge G) &= 1 + \max(d(F), d(G)) & D(F \wedge G) &= \max(D(F), D(G)) \\
d(\neg F) &= 1 + d(F) & D(\neg F) &= D(F) \\
d([a]F) &= 1 + d(F) & D([a]F) &= D(F) \\
d([A, e]F) &= 1 + d(F) & D([A, e]F) &= d(F)
\end{aligned}
$$

The number $d(F)$ tells us the maximum number of steps required to construct a part of F without counting recursive steps used to construct preconditions of action models. The number $D(F)$ tells us the maximum number of steps required to construct a part of a formula in F that comes just after an action, again without counting recursive steps used for precondition construction. (So in each case, a "step" is a recursive call to

the grammar (EAL^n) for some fixed n such that $F \in (\text{EAL}^n)$.) Proceeding, let $L(F)$ denote the *recursion level* of the (EAL)-formula F; that is, $L(F)$ is the smallest non-negative integer n such that $F \in (\text{EAL}^n)$. Finally, we order (EAL)-formulas as follows: assign to each (EAL)-formula F the pair $(L(F), D(F))$ and then order according to the strict dictionary ordering on pairs of integers (i.e., order first by the left number and then by the right number).

Dictionary Ordering of (EAL)-formulas. $F < G$ means that either $L(F) < L(G)$ or else both $L(F) = L(G)$ and $D(F) < D(G)$.

Note that we have $L(\mathsf{pre}^A(e)) < L([A, e]F)$ for each $(A, e) \in \mathsf{AM}_*$ and $F \in (\text{EAL})$. One can then check that the following holds.

EAL Dictionary Ordering Lemma. For each of the schemes 1–4 listed above, if F_1 denotes the formula on the left side of the equivalence and F_2 denotes the formula on the right, then $F_2 < F_1$.

Another useful ordering introduced by van Ditmarsch, van der Hoek, and Kooi (2007) is given by assigning a non-negative integer "complexity" $c(F)$ to each (EAL)-formula and then ordering according to the usual strict ordering on the integers.

$$
\begin{aligned}
c(p) &= 0 \text{ for letters } p \in \mathcal{P} \\
c(F \wedge G) &= 1 + \max(c(F), c(G)) \\
c(\neg F) &= 1 + c(F) \\
c([a]F) &= 1 + c(F) \\
c([A, e]F) &= (4 + c(A)) \cdot c(F) \\
c(A) &= \max\{c(\mathsf{pre}^A(f)) \mid f \text{ is an event in } A\}
\end{aligned}
$$

EAL Complexity Ordering Lemma (van Ditmarsch, van der Hoek, and Kooi 2007). For each of the schemes 1–4 listed above, if F_1

denotes the formula on the left side of the equivalence and F_2 denotes the formula on the right, then $c(F_2) < c(F_1)$.

The ordering lemmas make precise the idea that the right side of a given reduction axiom is "less complex" than the left side. These lemmas are used extensively to prove important results whose proof, like that of the Reduction Theorem, calls for an induction on (EAL)-formulas.

Finally, we note that showing the correctness of our definition of the binary satisfaction relation $\models$ between the class $\mathcal{C}_*$ of pointed Kripke models and the set of (EAL)-formulas can be done by way of an induction as well. In particular, we may use the ordering lemmas or the value n for the languages (EAL^n)—the latter is probably most natural—to provide an inductive construction of a binary satisfaction relation $\models\ \subseteq\ \mathcal{C}_* \times$ (EAL) between pointed Kripke models and (EAL)-formulas that minimally extends the binary satisfaction relation $\models\ \subseteq\ \mathcal{C}_* \times$ (ML) for the basic modal language (see Appendix A) such that for each $[A,e]G \in$ (EAL), we have the following:

- $M,w \models [A,e]G$ holds if and only if $M,w \not\models \mathsf{pre}^A(e)$ or $M[A],(w,e) \models G$, where the Kripke model $M[A]$ is defined via the BMS product update (Baltag, Moss, and Solecki 1998).

Correctness of the definition of satisfaction relations for other action model-style languages can be shown using similar methods.

Appendix I: Variants of the action model approach to Dynamic Epistemic Logic

1. Graph modifier logics

Aucher et al. (2009) study highly expressive extensions of (ML) that contain modalities for performing the following Kripke model-transforming operations:

- add a new world that is disconnected from all others and at which all propositional letters are false;
- add a new world as above but then consider this new world to be the "actual" one (i.e., consider it the new point of evaluation);
- given a formula F and a letter p, change the valuation of p by making p true only at those non-F worlds where it was true before;
- given a formula F and a letter p, change the valuation of p by making p true both where it was before and at all F-worlds;
- for an agent a and formulas G and G', remove all a-arrows in the model whose source is a G-world and whose target is a G'-world; and
- for an agent a and formulas G and G', add an a-arrow from every G-world to every G'-world.

Based on a language that extends (ML) with a universal modality, a modal operator for each of the operations listed above, and Propositional Dynamic Logic-style sequential combinations of these modalities (with test), Aucher et al. (2009) define a sound and complete logic GML of global graph modifiers. This logic extends public announcement logic. Aucher et al. (2009) also study a variant of GML that restricts the valuation-change operations so that only the valuation of the "actual world" (i.e., the point of evaluation) changes; this has natural connections with hybrid logic and is related to the work on local state assignment by Renardel de Lavalette (2004). Nevertheless, the exact connection between the Aucher et al. (2009) graph modification operations and action model-induced transformations is unknown.

Graph modifier logics may be considered part of the family of works that have grown out of van Benthem's (2005) "sabotage logic". Other examples in this family include van Eijck and Wang (2008), Areces, Fervari, and Hoffmann (2012), and Fervari (2014).

2. Generalized Arrow Update Logic

Generalized Arrow Update Logic (Kooi and Renne 2011b) is a variant of Arrow Update Logic (see Appendix E) that drops the assumption that arrow updates are common knowledge. This is achieved using a product update-like operation based on the basic arrow update language (AUL) (see Appendix E). In addition to providing a complete axiomatization (based on reduction axioms to (ML)), Kooi and Renne (2011b) study the *update expressivity* of generalized arrow updates with respect to action models. Note that this notion of *update expressivity* is very different from the notion of *language expressivity*. In particular, update expressivity is concerned with identifying the Kripke model-transforming functions can be described using the model-changing modalities of the language (e.g., the generalized arrow updates or the action models). Contrast this with language expressivity, which is concerned with identifying the propositions (here meaning "sets of worlds") that are definable using a formula of the language. Kooi and Renne (2011b) show that both the language (EAL) of Epistemic Action Logic (i.e., the logic of action models) and the language (GAUL) of Generalized Arrow Update Logic are equally language expressive and equally update expressive. However, (GAUL) may have some advantages:

- there exist generalized arrow updates whose smallest update-equivalent action models are exponentially larger; and
- generalized arrow updates are at worst poly-exponentially less succinct than action models, though this improves to being at worst polynomially less succinct if the action models have preconditions in (ML) or if generalized arrow updates are allowed to have target conditions in (EAL).

Therefore, if we restrict preconditions to (ML) or extend (GAUL) so as to allow (EAL)-formulas in target conditions—neither of which affects

update expressivity because (ML), (EAL), and (GAUL) are all equally language expressive—then generalized arrow updates are exponentially more succinct than action models.

Kooi and Renne (2011a) study a relativized common knowledge-style extension of the basic arrow update language (AUL); however, this extension has not yet been brought to the generalized language (GAUL). Finally, we mention that it seems as though the arrow-deletion operations of (AUL) ought to have some natural connection with the arrow-deletion operations in the graph modifier logics of Aucher et al. (2009). However, the connection is not obvious because arrow updates of (AUL) retain arrows that simultaneously satisfy a number of arrow specifications, whereas the graph modifier operations only call for satisfaction of what is essentially a single arrow specification. It may be possible to overcome this restriction, perhaps by addition of a Propositional Dynamic Logic-like union operation, but this has yet to be investigated.

3. Logic of Communication and Change

Van Benthem, van Eijck, and Kooi (2006) introduced LCC, the Logic of Communication and Change, as a Propositional Dynamic Logic-like language that incorporates action models with substitution (see Appendix G). The language (LCC) of the Logic of Communication and Change and the set $\mathsf{AM}(\mathsf{LCC})_*$ of pointed action models with substitution having preconditions in the language (LCC) are together defined by the following recursive grammar:

$$\begin{aligned} F &::= p \mid F \wedge F \mid \neg F \mid [\pi]F \mid [A,e]F \\ \pi &::= a \mid ?F \mid \pi;\pi \mid \pi \cup \pi \mid \pi * \\ &\quad p \in \mathcal{P},\ a \in \mathcal{A},\ (A,e) \in \mathsf{AM}(\mathsf{LCC})_* \end{aligned} \tag{LCC}$$

Expressions made using the start symbol F are called *formulas*, and expressions made using the start symbol π are called *programs*. The sublanguage (EPDL) of “Epistemic Propositional Dynamic Logic” is defined by disallowing formation of action model formulas $[A, e]F$ in the grammar for (LCC). The semantics of this language is given as follows.

- $M, w \vDash p$ holds if and only if $w \in V(p)$.
- $M, w \vDash F \wedge G$ holds if and only if $M, w \vDash F$ and $M, w \vDash G$.
- $M, w \vDash \neg F$ holds if and only if $M, w \nvDash F$.
- $M, w \vDash [\pi]F$ holds if and only if $M, v \vDash F$ for each v such that $w[\![\pi]\!]v$.
- $M, w \vDash [A, e]F$ holds if and only if $M, w \nvDash \mathsf{pre}^A(e)$ or $M[A], (w, e) \vDash F$, where $M[A]$ is defined by the BMS product update (Baltag, Moss, and Solecki 1998).
- $x[\![a]\!]y$ holds if and only if $xR_a y$.
- $x[\![?F]\!]y$ holds if and only if $x = y$ and $M, w \vDash F$.
- $x[\![\pi_1 ; \pi_2]\!]y$ holds if and only if there exists z such that $x[\![\pi_1]\!]z$ and $z[\![\pi_2]\!]y$.
- $x[\![\pi_1 \cup \pi_2]\!]y$ holds if and only if $x[\![\pi_1]\!]y$ or $x[\![\pi_2]\!]y$.
- $x[\![\pi*]\!]y$ holds if and only if $x([\![\pi]\!]^*)y$ holds, where $[\![\pi]\!]^*$ is the reflexive-transitive closure of $[\![\pi]\!]$.

Van Benthem, van Eijck, and Kooi (2006) provide a sound and complete axiomatization for the logic LCC consisting of the validities of (LCC). Completeness is via reduction to (EPDL). The authors also show that the language is highly expressive, including in particular relativized common knowledge among a group $\{a_1, \dots, a_n\}$ via the program

$$(a_1 \cup \cdots \cup a_n)*.$$

We refer the reader to van Benthem, van Eijck, and Kooi (2006) for details.

Van Eijck and Wang (2008) consider an extension of LCC that allows converses of agent programs a and apply this to consensus seeking in

Dutch plenary meetings. Van Eijck (2008b) and van Eijck and Wang (2008) consider a further extension that allows for *relational change* operators; these are similar to valuation changes except that programs are used to describe Kripke model transformations that include changes in agents’ relations R_a in the model. These extensions, which are extremely expressive, are applied to the study of Belief Revision in a DEL-style framework.

4. General Dynamic Dynamic Logic

Girard, Seligman, and Liu (2012) propose General Dynamic Dynamic Logic GDDL, a Propositional Dynamic Logic-style language that has complex action model-like modalities that themselves contain Propositional Dynamic Logic-style instructions. The basic idea, following van Benthem and Liu (2007) and Liu (2008), is to describe relation changes using programs but then use additional “higher level” programs to describe how the “lower level” programs are to interact. The framework is mathematically complicated, so we refer the reader to Girard, Seligman, and Liu (2012) for details. We note that Girard, Seligman, and Liu (2012) argue that GDDL has model-transforming updates that are sufficiently expressive to express all model-changing transformations of action models, Generalized Arrow Update Logic, or LCC. However, it has not been proved whether this result is strict (i.e., whether all GDDL-transformations are expressible using, e.g., LCC). Seligman, Liu, and Girard (2011, 2013) apply GDDL to a logic of “friendship” (a certain notion of social connectivity) in social networks.

Appendix J: Conditional Doxastic Logic

The language (CDL) of *conditional doxastic logic* is defined by the following grammar:

$$F ::= p \mid F \wedge F \mid \neg F \mid B_a^F F \qquad \text{(CDL)}$$
$$p \in \mathcal{P},\ a \in \mathcal{A}$$

We think of the formula $B_a^F G$ as saying that agent a believes F after conditionalization (or static belief revision) by G. The satisfaction relation $\models$ between pointed plausibility models and formulas of (CDL) and the set $[\![F]\!]_M$ of worlds in the plausibility model M at which formula F is true are defined by the following recursion.

- $[\![F]\!]_M := \{w \in W \mid M, w \models F\}$.
- $M, w \models p$ holds if and only if $w \in V(p)$.
- $M, w \models F \wedge G$ holds if and only if $M, w \models F$ and $M, w \models G$.
- $M, w \models \neg F$ holds if and only if $M, w \not\models F$.
- $M, w \models B_a^G F$ holds if and only if $\min_a([\![G]\!]_M \cap \mathsf{cc}_a(w)) \subseteq [\![F]\!]_M$. $B_a^G F$ means that "F is true at the most plausible G-worlds consistent with a's information".

In the language (CDL), the operator K_a for possession of information may be defined as follows:

$$K_a F \text{ denotes } B_a^{\neg F}\bot,$$

where $\bot$ is the propositional constant for falsehood. We then have the following.

> **Theorem (Baltag and Smets 2008b).** For each pointed plausibility model (M, w), we have:
>
> $$M, w \models B_a^{\neg F}\bot \text{ iff } M, v \models F \text{ for each } v \text{ with } v \simeq_a w.$$
>
> So the (CDL)-abbreviation of $K_a F$ as $B_a^{\neg F}\bot$ gives the same meaning in (CDL) for "possession of information" as we had in ($K\Box$).

The validities of (CDL) are axiomatized as follows.

The axiomatic theory CDL **(Baltag, Renne, and Smets 2015).**

- Axiom schemes and rules for classical propositional logic
- $B_a^F(G_1 \to G_2) \to (B_a^F G_1 \to B_a^F G_2)$
- $B_a^F F$
- $B_a^F \bot \to B_a^{F\wedge G}\bot$
- $\neg B_a^F \neg G \to (B_a^F H \to B_a^{F\wedge G} H)$
- $B_a^{F\wedge G} H \to B_a^F(G \to H)$
- $B_a^{F\wedge G} H \to B_a^{G\wedge F} H$
- $B_a^F H \to B_a^G B_a^F H$
- $\neg B_a^F H \to B_a^G \neg B_a^F H$
- $B_a^F \bot \to \neg F$

CDL **Soundness and Completeness (Board 2004; Baltag and Smets 2008b; Baltag, Renne, and Smets 2015).** CDL is sound and complete with respect to the collection $\mathcal{C}_*$ of pointed plausibility models. That is, for each (CDL)-formula F, we have $\mathsf{CDL} \vdash F$ if and only if $\mathcal{C}_* \models F$.

Appendix K: Evidential dynamics and justified belief

In this Appendix, we examine work in DEL aimed at reasoning about evidence, belief, and knowledge.

- 1. Connections with Justification Logic
- 2. Connections with neighborhood models

---

1. Connections with Justification Logic

In this section, we overview the Baltag, Renne, and Smets (2014) study of a logic of "conclusive" (or "good") evidence based on a combination of plausibility models with an adaptation of the syntactic bookkeeping mechanisms of Justification Logic (Artemov 2008; see entry on Justification Logic). The language (JBG) of the logic JBG of *Justified Belief with Good Evidence* is an extension of the language ($K\Box$) obtained by adding a few additional features. First, (JBG) introduces structured objects t called *terms* that encode evidence, reasons, or justifications (words we use synonymously here). Second, (JBG) adds to ($K\Box$) two new kinds of formulas for reasoning about terms: $E_a t$ ("agent a possesses evidence t") and $t \gg F$ ("t is admissible (as evidence) for F").

$$\begin{aligned} F &::= p \mid F \wedge F \mid \neg F \mid \Box_a F \mid K_a F \mid E_a t \mid t \gg F \\ t &::= c_F \mid t \cdot t \mid t + t \\ & p \in \mathcal{P},\ a \in \mathcal{A} \end{aligned} \tag{JBG}$$

Objects formed using the symbol F in the above grammar are called *formulas*. Objects formed using the symbol t in the above grammar are called *terms*; a term of the form c_F is called a *certificate for F*. The dual $\Diamond_a$ of $\Box_a$ is defined by $\Diamond_a F := \neg\Box_a\neg F$ and the dual $\hat{K}_a$ of K_a is defined by $\hat{K}_a F := \neg K_a \neg F$. The set $\mathsf{sub}(t)$ of *subterms* of a term t is defined by induction on the construction of t as follows:

$$\begin{aligned} \mathsf{sub}(c_F) &= \{c_F\}, \\ \mathsf{sub}(t \cdot s) &= \{t \cdot s\} \cup \mathsf{sub}(t) \cup \mathsf{sub}(s), \text{ and} \\ \mathsf{sub}(t + s) &= \{t + s\} \cup \mathsf{sub}(t) \cup \mathsf{sub}(s). \end{aligned}$$

Terms are used to represent evidence, reasons, or justification in support of a formula. A number of evidential relationships between a term and a formula may be introduced. Each depends on two key components. First, the term must be "admissible" for the formula, a technical notion defined in a moment. Roughly speaking, to say a term is admissible for a formula means that the term has the proper "shape" of a logical argument for that formula. Second, the certificates that make up the term must be "conclusive" (or "good"). This means that a formula F named in any certificate c_F occurring in the term must be true. Let us first look at admissibility.

Admissibility. *Admissibility* is the smallest binary relation $\gg'$ between terms and formulas satisfying the following:

- $c_F \gg' F$,
- $t \gg' (F \to G)$ and $s \gg' F$ together imply $(t \cdot s) \gg' G$,
- $t \gg' F$ and $s \gg' G$ together imply $(t + s) \gg' F$, and
- $t \gg' F$ and $s \gg' G$ together imply $(s + t) \gg' F$.

In the forthcoming semantics, we will see that the formula $t \gg F$ is true if and only if $t \gg' F$ holds. So we may sometimes conflate $\gg'$ and $\gg$ without risking confusion. Further, we see that admissibility does describe terms as having a proper logical "shape":

- a certificate c_F is admissible for the formula F it certifies and so certificates may be used to represent evidence for observations (including introspection as self-observation) or for other basic evidenced assumptions;
- the (non-commutative) term-combining operator "$\cdot$" combines terms so that admissibility follows via the logical rule of *Modus Ponens*, thereby encoding one step of logical derivation; and
- the (commutative) term-combining operator "$+$" combines terms so that admissibility follows via simple aggregation without logical inference, thereby encoding the "grouping together" of separate pieces of evidence into a unified whole.

The condition on the formula G in the definition of admissibility for the "$+$" operation is there to ensure that a term $t + s$ that is admissible for some formula (and "well-shaped") satisfies the property that each of its component terms is also admissible for some formula (and is also "well-shaped"). We note that the *content* of a term t, defined as the set

$$\mathsf{con}(t) := \{\theta \mid t \gg \theta\}$$

of formulas for which t is admissible, is finite and computable. In particular, by examining the stepwise construction of a term according to the grammar defining terms, we see that each inductive step in the construction specifies either a *Modus Ponens* operation "$\cdot$" or an aggregation operation "$+$". This enables us to establish a "chain of evidence" that begins with certificates and proceeds stepwise via *Modus Ponens* or aggregation until the full term is constructed and a particular admissible formula F is found.

The formula $E_a t$ says that agent a possesses the term t. This means that a has done all the work necessary to collect, examine, and make proper sense of the term t. The set of terms that the agent possesses at world w is denoted by $E_a(w)$.

To define the notion of evidential "conclusiveness" (or "goodness"), we must define the models for (JBG) in two stages. In the first, we define "plausibility evidence pre-models", which are just plausibility models to which we add sets $E_a(w)$ naming the evidence in possession for each agent a and world w.

Plausibility evidence pre-model. Given a nonempty set $\mathcal{P}$ of propositional letters and a finite nonempty set $\mathcal{A}$ of agents, an *plausibility evidence pre-model* is a structure

$$M = (W, \geq, E, V)$$

for which $(W, \geq, V)$ is a plausibility model and E is a function that assigns to each agent a and each world w a set $E_a(w)$ of terms such that the following are satisfied:

- Trivial Evidence: $c_\top \in E_a(w)$.

- Certification of Evidence: if $t \in E_a(w)$ and $t \gg F$, then $c_F \in E_a(w)$.
- Subterm Closure: if $t \in E_a(w)$, then $\mathsf{sub}(t) \subseteq E_a(w)$.
- Availability of Evidence:
  - if $t \cdot s \in E_a(w)$, $w' \in \mathsf{cc}_a(w)$, and $\{t, s\} \subseteq E_a(w')$, then $t \cdot s \in E_a(w')$; and
  - if $t + s \in E_a(w)$, $w' \in \mathsf{cc}_a(w)$, and $\{t, s\} \subseteq E_a(w')$, then $t + s \in E_a(w')$.

A *pointed plausibility evidence pre-model* is a pair (M, w) consisting of a plausibility evidence pre-model M and a world w (called the *point*) that designates the state of affairs that we (the formal modelers) currently assume to be actual.

Trivial Evidence says that each agent always has evidence for *something*, even if this is only the trivial the propositional constant $\top$ for truth. Certification of Evidence says that if the agent has t and this term is admissible for F, then the agent also has the certificate c_F for F. This tells us that agents can always "convert" complicated pieces of evidence into certificates, thereby ignoring the complicated term structure and keeping track only of the fact that there was *some* evidence in support of F. Subterm closure says that an agent cannot possess a piece of evidence without also possessing its constituent parts. Availability of Evidence says that whenever the agent has constructed a piece of evidence t and there is an informationally consistent world w' at which she has also constructed the components that make up t, then she has also constructed t at w'. Intuitively, this says that the agent has information about the evidence she has constructed.

The satisfaction relation $\models$ between pointed plausibility evidence pre-models and formulas of (JBG) is defined as follows.

- $M, w \models p$ holds if and only if $w \in V(p)$.
- $M, w \models F \wedge G$ holds if and only if both $M, w \models F$ and $M, w \models G$.
- $M, w \models \neg F$ holds if and only if $M, w \not\models F$.
- $M, w \models \Box_a F$ holes if and only if $M, v \models F$ for each $v \leq_a w$.
- $M, w \models K_a F$ holds if and only if $M, v \models F$ for each $v \in \mathsf{cc}_a(w)$.
- $M, w \models E_a t$ holds if and only if $t \in E_a(w)$.
- $M, w \models t \gg F$ holds if and only if $t \gg' F$.

The models of (JBG) are the pre-models in which evidence is "conclusive" (or, synonymously, "good"). This means that available certificates are fully reliable: what they certify is true.

> **Plausibility evidence model.** Given a nonempty set $\mathcal{P}$ of propositional letters and a finite nonempty set $\mathcal{A}$ of agents, a *plausibility evidence model* is a plausibility evidence pre-model M (based on $\mathcal{P}$ and $\mathcal{A}$) satisfying the following property for each $a \in \mathcal{A}$ and $F \in$ (JBG):
>
> - Evidential Goodness: $c_F \in E_a(w)$ implies $M, w \models F$.
>
> A *pointed plausibility evidence model*, sometimes called a *scenario* or a *situation*, is a pair (M, w) consisting of a plausibility evidence model M and a world w (called the *point*) that designates the state of affairs that we (the formal modelers) currently assume to be actual.

Because the evidence-combining operations "$\cdot$;" of *Modus Ponens* and "$+$" of aggregation are logically sound, it is not difficult to see that *all* evidence in possession is "good" (i.e., "conclusive") in plausibility evidence models; that is, for a plausibility evidence model M, if we have $t \in E_a(w)$ and $t \gg F$, then $M, w \models F$. Also, while Evidential Goodness may at first seem too strong (in that it postulates infallibility of available evidence), we note that agents still can have a fallible relationship with respect to evidence even with Evidential Goodness in place. In particular,

it is possible for agent a to believe she possesses evidence when she in fact does not (i.e., $E_a t$ is true at the most plausible worlds but false at the actual world), and it is also possible for agent a to believe that she does not possess evidence that she in fact does (i.e., $E_a t$ is false in some most plausible world but true at the actual world). So we see that Evidential Goodness still allows evidence to be misleading.

The following are abbreviations defined notions built upon the simple language (JBG):

- Unconditional belief: $B_a F = \Diamond_a \Box_a F$.
- Conditional belief: $B_a^G F = \hat{K}_a G \to \hat{K}_a (G \wedge \Box_a (G \to F))$.
- Implicit justification: $t{:}F = (t \gg F) \wedge \bigwedge \{\theta \mid c_\theta \in \mathsf{sub}(t)\}$.
- Explicit justification: $t{:}^e_a F = E_a t \wedge (t \gg F)$.
- Explicit belief: $B^e_a F = B_a E_a c_F$.
- Explicit conditional belief: $B^{e,G}_a F = B^G_a E_a c_F$.
- Explicit defeasible knowledge: $\Box^e_a F = \Box_a E_a c_F$.
- Explicit information possession: $K^e_a F = K_a E_a c_F$.

Conditional and unconditional belief are equivalent to the definition of conditional belief and unconditional belief for the language ($K\Box$). The other notions are new.

- *Implicit Justification*: to say that t is implicit justification for F, written $t{:}F$, means that t is admissible for F and the certificates that make up t are "good" (i.e., what they certify is true). Intuitively, this says that t is a piece of evidence that objectively supports F, no matter whether any agent actually has t in her possession. It is in this sense that the justification is "implicit" as opposed to "explicit", which would require some agent to have gone to the trouble to gather, process, and make sense of t. Implicit justification is closely related to some common schematic principles from Justification Logic; see

Baltag, Renne, and Smets (2014) for details.

- *Explicit Justification*: to say that agent a has explicit justification for F, written $t:^e_a F$, means that t is implicit justification for F and a actually has t in her possession. If this is so then, intuitively, a has gone to the trouble to gather, process, and make sense of t, and therefore a explicitly possesses this evidence for F.
- *Other explicit notions*: the notions of explicit belief, explicit conditional belief, explicit defeasible knowledge, and explicit information concern the agent's propositional attitude with respect to the certificate c_F for a formula F: she believes she has the certificate, she conditionally believes she has it, she defeasibly knows she has it, or she has information that she has it.

The theory JBG of *Justified Belief with Good Evidence* is defined as follows.

The axiomatic theory JBG.

- Classical Logic: Axiom schemes and rules for classical propositional logic
- Admissibility:

$$t \gg F \text{ for each } t \text{ and } F \text{ such that } t \gg' F$$
$$\neg(t \gg F) \text{ for each } t \text{ and } F \text{ such that not } t \gg' F$$

- Trivial Evidence: $E_a c_\top$
- Certification of Evidence: $(E_a t \wedge t \gg F) \to E_a c_F$
- Subterm Closure: $(E_a(t \cdot s) \vee E_a(t + s)) \to (E_a t \wedge E_a s)$
- Availability of Evidence:

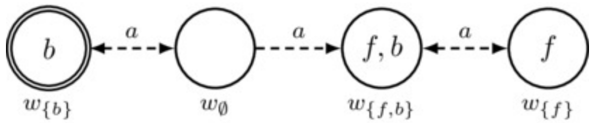
$$E_a(t \cdot s) \to K_a((E_a t \wedge E_a s) \to E_a(t \cdot s))$$
$$E_a(t + s) \to K_a((E_a t \wedge E_a s) \to E_a(t + s))$$

- Evidential Goodness: $E_a c_F \to F$
- Information: S5 axiom schemes and rules for K_a
- Defeasible Knowledge: S4 axiom schemes and rules for $\Box_a$
- Indefeasibility: $K_a F \to \Box_a F$
- Local Connectedness: $K_a(\Box_a F \to G) \vee K_a(\Box_a G \to F)$

Soundness, Completeness, and Finite Model Property. To say an evidence model is *finite* means the set of worlds is finite and each set $E_a(w)$ is also finite. JBG is sound and complete with respect to the collection $\mathcal{C}_*$ of pointed finite evidence models. That is, for each (JBG)-formula F, we have that JBG $\vdash F$ if and only if $\mathcal{C}_* \models F$.

JBG can be used to reason about certain examples from Epistemology. For example, Gettier (1963) constructs a famous counterexample to the claim that "knowledge" may be equated with "justified true belief" (i.e., justified *correct* belief). In this example, an agent—let us call her a—has evidence for a propositional letter f, concludes via logical deduction that $b \vee f$, and therefore has evidence for this disjunction; however, unknown to the agent, f is false but b is true. She therefore has justified true belief but not knowledge that $b \vee f$ is true (since her reason for believing this disjunction is based on her belief in the wrong disjunct). This example is easily reconstructed in JBG using the pointed plausibility evidence model in Figure K1. In this figure, the f-worlds are more plausible than the $\neg f$ worlds (since the agent believes f) and, within each group of equi-plausible worlds, there is both a b-world and a $\neg b$-world (representing the fact that she has no information about b). In the actual world, b is true but f is false. The agent actually has available to her only the evidence $c_{f\to(b\vee f)}$ representing her knowledge of the logical truth $f \to (b \vee f)$. However, the agent mistakenly believes she has available to her the good evidence c_f and the combined evidence $c_{f\to(b\vee f)} \cdot c_f$ for $b \vee f$ (along with the certificate $c_{b\vee f}$ in accordance with Certification of Evidence). Also, by Trivial Evidence, she has (and believes she has) the trivial certificate $c_\top$. Therefore, as is readily verified, the agent has a justified true belief of

$b \vee f$ based on the evidence $c_{f \to (b \vee f)} \cdot c_f$, which explicitly represents her reasoning, and yet she does not have defeasible knowledge of $b \vee f$ because she is mistaken about her evidence.



$$
\begin{aligned}
E_a(w_{\{b\}}) &:= \{c_\top, c_{f \to b \vee f}\} \\
E_a(w_\emptyset) &:= E_a(w_{\{b\}}) \\
E_a(w_{\{f,b\}}) &:= \{c_\top, c_f, c_{f \to b \vee f}, c_{f \to b \vee f} \cdot c_f, c_{b \vee f}\} \\
E_a(w_{\{f\}}) &:= E(w_{\{f,b\}})
\end{aligned}
$$

$$M_G$$

Figure K1. The Gettier Example: justified true belief without knowledge; that is, we have $M_G, w_{\{b\}} \models (b \vee f) \wedge B_a^e(b \vee f)$ and yet we have $M_G, w_{\{b\}} \not\models \Box_a^e(b \vee f)$ (i.e., no explicit knowledge) and $M_G, w_{\{b\}} \not\models \Box_a(b \vee f)$ (i.e., no implicit knowledge).

Baltag, Renne, and Smets (2014) also study another example of "explicit justified true belief and implicit knowledge without explicit knowledge" and discuss how JBG may be used to avoid the problem of (logical) omniscience, among other matters. A version of JBG with DEL-style dynamics (of addition of new terms, dynamic *Modus Ponens*, announcement-like addition of evidence with world elimination, and evidential upgrade) is studied in Baltag, Renne, and Smets (2012). The latter is part of what we might call *Dynamic Justification Logic*, the joint literature on Justification Logic and Dynamic Epistemic Logic. Other works in this joint field include work by Renne on evidence elimination (Renne 2009, 2011b, 2012) and on public announcements in Justification Logic-style languages (Renne 2011a), by Bucheli et al. (2010) on

Gerbrandy–Groeneveld-style public announcements in a Justification Logic setting, and by Bucheli, Kuznets, and Studer (2011) on the "Realization Problem" (concerning the relationship between theorems of a Justification Logic-based theory with syntactically similar theorems of a modal logic-based theory) for a logic with GG-style announcements.

Finally, a discussion in Baltag, Renne, and Smets (2012) describes how the work there may be viewed as a generalization and refinement of other DEL-style logics of inference and update (Velázquez-Quesada 2009; van Benthem and Velázquez-Quesada 2010).

2. Connections with neighborhood models

A different approach to evidence in Dynamic Epistemic Logic was proposed by van Benthem and Pacuit (2011a,b) and studied further in van Benthem Fernández-Dunque, and Pacuit (2012, 2014). This approach is much less syntactic than the Justification Logic-style approaches, focusing instead on the semantic notion of modal "neighborhood" (or "minimal") models that have been repurposed with an evidential twist.

Neighborhood evidence models. Given a nonempty set $\mathcal{P}$ of propositional letters and a finite nonempty set $\mathcal{A}$ of agents, an *neighborhood evidence model* is a structure

$$M = (W, E, V)$$

consisting of

- a nonempty set W of *worlds* identifying the possible states of affairs that might obtain;
- a function $E : W \to \wp(\wp(W))$ that maps each agent a and each world w to a set $E_a(w)$ of sets of worlds such that
  - $\emptyset \notin E_a(w)$ and

- $W \in E_a(w)$; and
- a *propositional valuation* $V : \mathcal{P} \to \wp(W)$ mapping each propositional letter to the set of worlds at which that letter is true.

A *pointed neighborhood evidence model*, sometimes called a *scenario* or a *situation*, is a pair (M, w) consisting of a neighborhood evidence model M and a world w (called the *point*) that designates the state of affairs that we (the formal modelers) currently assume to be actual.

Let us call a set of worlds a *proposition*. Intuitively, a proposition is true at just those worlds it contains. The set $E_a(w)$ represents the propositions for which agent a has evidence at world w. The condition $\emptyset \notin E_a(w)$ ensures that the agent does not have evidence for a contradiction (i.e., the proposition $\emptyset$, which is true at no worlds). The condition $W \in E_a(w)$ says that the agent has some evidence at her disposal, even if it is only for the trivial proposition consisting of all worlds.

An agent may have inconsistent evidence at a world. For example, the agent might have $P \in E_a(w)$ and $Q \in E_a(W)$ for disjoint nonempty propositions P and Q. As van Benthem and Pacuit (2011a,b) observe, such situations lead to a number of possibilities for defining agent belief. One that van Benthem and Pacuit (2011a,b) and van Benthem, Fernández-Dunque, and Pacuit (2012, 2014) study in detail goes as follows: the agent should combine as much mutually consistent evidence as is possible, and she is said to believe whatever is true at all of the worlds in this maximally consistent combination. Here the natural notion of evidence combination is intersection: if propositions P and P' are *consistent*, meaning they have nonempty intersection, then their combination is just the intersection $P \cap P'$ consisting of the worlds supported by both propositions. So to combine as much consistent evidence as is possible, the agent should intersect as many propositions from $E_a(w)$ as possible such that the empty set is not the result. There may be multiple different ways to do this, some

yielding different final nonempty intersections. Nevertheless, once she has reached any such maximal nonempty intersection, then the formulas true at all worlds in the intersection are those the agent believes. Or, put in the language of propositions, the agent believes all propositions that are extensions (i.e., supersets) of some maximal nonempty intersection of evidenced propositions. A maximal nonempty intersections of evidenced propositions is called a "scenario" in van Benthem, Fernández-Dunque, and Pacuit (2012). To avoid a conflict with other uses of the word "scenario" in this article, we will use different terminology.

Body of evidence. Let (M, w) be a pointed neighborhood evidence model. A *body of evidence* at world w is a nonempty subset $S \subseteq E_a(w)$ of propositions. When convenient, a singleton body of evidence $\{P\}$ may be identified with the proposition P making up the singleton set. To say a body of evidence S is *consistent* means that $\bigcap S \neq \emptyset$; an *inconsistent* body of evidence is one that is not consistent. To say that a body of evidence S at w is *maximal consistent* means that S is consistent and any body of evidence T at w that strictly contains S (i.e., $T \supsetneq S$) is inconsistent. A *belief set* at w is the intersection of a maximal consistent body of evidence at w.

Van Benthem and Pacuit (2011a,b) and van Benthem, Fernández-Dunque, and Pacuit (2012) consider the following language, which we call (NEL), for reasoning about neighborhood evidence models. (A wider variety of languages is studied in van Benthem, Fernández-Dunque, and Pacuit 2014, but we focus on this language here.) The language (NEL) uses doxastic modal formulas $B_a F$ ("agent a believes F"), modal formulas $\Box_a F$ ("agent a has evidence for F"), and a universal modality $[\forall]$.

$$F ::= p \mid F \wedge F \mid \neg F \mid B_a F \mid \Box_a F \mid [\forall]F \qquad \text{(NEL)}$$
$$p \in \mathcal{P},\ a \in \mathcal{A}$$

The satisfaction relation $\models$ between pointed neighborhood evidence models and formulas of (NEL) is defined as follows.

- $M, w \models p$ holds if and only if $w \in V(p)$.
- $M, w \models F \wedge G$ holds if and only if both $M, w \models F$ and $M, w \models G$.
- $M, w \models \neg F$ holds if and only if $M, w \not\models F$.
- $M, w \models B_a F$ holds if and only if there exists a belief set P at w such that $M, v \models F$ for each $v \in P$.
- $M, w \models \Box_a F$ holds if and only if there exists $P \in E_a(w)$ such that $M, v \models F$ for each $v \in P$.
- $M, w \models [\forall] F$ holds if and only if $M, v \models F$ for each $v \in W$.

Van Benthem and Pacuit (2011a,b) also define extensions of (NEL) with conditional evidential existence and conditional belief. The conditional evidential existence formula $\Box_a^G F$ is true at a world w if and only if there is an evidenced proposition $P \in E_a(w)$ that is consistent with the set $[\![G]\!]$ of G-worlds and that is contained in the set $[\![F]\!]$ of F-worlds. The conditional belief formula $B_a^G F$ is true at a world w if and only if there exists a belief set P at w that is consistent with $[\![G]\!]$ and contained in $[\![F]\!]$. These conditional versions of evidence existence and belief are used in the study of a number of different DEL-style dynamic operations on neighborhood evidence models (van Benthem and Pacuit 2011a,b).

- Public announcement of F: this is the straightforward adaptation of public announcements to the neighborhood models setting. In particular, the public announcement of F performs the operation

  $$(W, E, V) \mapsto (W^{F!}, E^{F!}, V^{F!})$$

  defined by setting
  - $W^{F!} := W \cap [\![F]\!]$,
  - $E_a^{F!}(w) := \{P \cap W^{F!} \mid P \in E_a(W)\}$, and
  - $V^{F!}(p) := V(p) \cap W^{F!}$.

- Addition of evidence for F: this adds evidence for F at each world according to the operation

  $$(W, E, V) \mapsto (W, E^{+F}, V)$$

  defined by setting

  $$E_a^{+F}(w) := E_a(w) \cup \{[\![F]\!]\}.$$

- Removal of evidence for F: this removes evidence at each world for anything that implies F via the operation

  $$(W, E, V) \mapsto (W, E^{-F}, V)$$

  defined by setting

  $$E_a^{-F}(w) := E_a(w) - \{P \in E_a(w) \mid P \subseteq [\![F]\!]\}.$$

- Upgrade of evidence for F: this makes evidence for F “important” by making sure every piece of evidence supports F. This is achieved by way of the operation

  $$(W, E, V) \mapsto (W, E^{\Uparrow F}, V)$$

  defined by setting

  $$E_a^{\Uparrow F}(w) := \{[\![F]\!]\} \cup \{P \cup [\![F]\!] \mid P \in E_a(w)\}.$$

- Evidence combination: this combines evidenced propositions via intersection according to the operation

  $$(W, E, V) \mapsto (W, E^{\sharp}, V)$$

  defined by setting

$$E_a^\sharp(w) := \{\bigcap S \mid S \text{ is a consistent body of evidence at } w\}.$$

Neighborhood evidence models and plausibility models are related, with neighborhood evidence models acting as a generalization of plausibility models, in the sense of the following theorem.

Theorem (van Benthem and Pacuit 2011a,b).

- Given a plausibility model $M = (W, \geq, V)$, define the neighborhood evidence model $\mathbb{m}(M) := (W, E, V)$ by setting

$$E_a(w) := \{P \subseteq W \mid P \neq \emptyset \text{ and } \forall x, \forall y : (x \in P \wedge y \leq_a x) \Rightarrow y \in P\}.$$

  That is, $E_a(w)$ consists of all nonempty propositions that are "closed" under increasing plausibility (i.e., if the proposition contains world x and world y is more plausible in the eyes of a than is x, then the proposition also contains y). Define the translation

$$(-)' : (\mathrm{NEL}) \to (K\Box{+}\forall)$$

  from (NEL)-formulas F to $(K\Box)$-formulas F' with a universal modality as follows: F' is obtained from the (NEL)-formula F by replacing every subformula $\Box_a G$ with $[\exists]\Box_a G'$ and replacing every subformula $B_a G$ with $[\forall]\Diamond_a[\exists]\Box_a G'$. For each (NEL)-formula F, we have

$$\mathbb{m}(M), w \vDash F \text{ iff } M, w \vDash F'.$$

  This result can be extended to conditional evidence possession and conditional belief as well (see van Benthem and Pacuit 2011b for details).
- Let $M = (W, E, V)$ be a neighborhood evidence model. If M is

  uniform, meaning $E_a(x) = E_a(y)$ for all worlds x and y, then we define the plausibility model $\mathbb{p}(M) = (W, \geq, V)$ as follows:

$$x \geq_a y \text{ means } \forall P \in E_a(w) : x \in P \text{ implies } y \in P.$$

  We have the following:
  - $\mathbb{p}(\mathbb{m}(M)) = M$ for each plausibility model M.
  - $\mathbb{m}(\mathbb{p}(M)) = M^\sharp$ for each neighborhood evidence model M.
  - The equality $\mathbb{m}(\mathbb{p}(M)) = M$ does not hold for every neighborhood evidence model M.

The neighborhood approach to evidence has not yet been applied to concrete examples from Epistemology, though perhaps this approach could provide a natural semantic backdrop for a Justification Logic-style approach such as that presented above. There are also possible connections with other Dynamic Epistemic Logic-style of logics of awareness and inference (van Benthem 2008b,c; Velázquez-Quesada 2009; van Benthem and Velázquez-Quesada 2010). Regardless, the neighborhood approach does provide natural and elegant evidence dynamics worth further study. We refer the reader to van Benthem and Pacuit (2011a,b) and van Benthem, Fernández-Dunque, and Pacuit (2012) for further details. Further study of additional structure on neighborhood models and other languages may be found in van Benthem, Fernández-Dunque, and Pacuit (2014).

Appendix L: Probabilistic update in Dynamic Epistemic Logic

Dynamic Epistemic Logics that incorporate probability have been studied by a number of authors. Van Benthem (2003), Kooi (2003), Baltag and Smets (2008a), and van Benthem, Gerbrandy, and Kooi (2009b) studied logics of finite probability spaces. Sack (2009) extended the work of Kooi

(2003) and van Benthem, Gerbrandy, and Kooi (2009b) to full probability spaces (based on σ-algebras of events). For a broader perspective on connections between logic and probability, see the *Stanford Encyclopedia of Philosophy* entry Logic and Probability (Demey, Kooi, and Sack 2013). In this appendix, we focus on two approaches in DEL: the conditional probability model approach of Baltag and Smets (2008a) and the epistemic probability model approach of van Benthem, Gerbrandy, and Kooi (2009b).

- 1. Conditional probabilities and multi-agent probabilistic update
- 2. Probabilistic Dynamic Epistemic Logic

---

1. Conditional probabilities and multi-agent probabilistic update

Our focus here will be on the case of finitely many probabilistic outcomes and on logics that connect three areas of work: the Popper–Réyni–de Finetti extension of Bayesian probabilistic conditionalization, the theory of Belief Revision, and Dynamic Epistemic Logic.

Five key design principles for a DEL-style probabilistic theory of Belief Revision guide how we proceed:

1. Belief should be equated with subjective probability that is at or above a fixed nonzero threshold.
2. Belief should be closed under finite conjunctions.
3. Beliefs can be false.
4. Static belief change should proceed by probabilistic conditionalization.
5. Dynamic belief change should proceed by a DEL-style probabilistic product update.

The meaning of the first principle is that we are to fix a nonzero belief threshold c in the real-number unit interval $(0,1]$ and identify the statement that "agent a believes F" with the statement that "agent a assigns probability $P_a(F) \geq c$ to the event that F is true". (A variant would be to take the strict inequality $P_a(F) > c$ instead, but we do not consider this variant here.) The second principle means that if an agent believes each of a finite number of statements $F_1, \dots, F_n$, then she should also believe the conjunction $F_1 \wedge \cdots \wedge F_n$. Taken together, these two principles require us to choose the threshold $c = 1$. The argument for this goes by way of the "Lottery Paradox" (van Fraassen 1995): for a fixed nonzero belief threshold $c \in (0,1)$ and an integer $n \geq \frac{1}{1-c}$, the agent will believe that any one of the tickets in a fair n-ticket lottery is not a winner (because she assigns a probability $1 - \frac{1}{n} \geq c$ to the statement that ticket i is not a winner). So she believes that ticket 1 is not a winner, that ticket 2 is not a winner, and so on through ticket n. It follows by closure under finite conjunctions that she must believe that *no ticket is a winner*. However, she assigns probability $1 \geq c$ to the event that *some ticket is a winner*, and therefore she also believes that *some ticket is a winner*—a contradiction. To avoid this contradiction and retain the first two design principles, we must choose the threshold $c = 1$ for belief: to say that "agent a believes F" means that agent a assigns probability 1 to the event the F is true. Written in a short motto, "belief is probability 1".

The third design principle reflects the fact that we are studying belief, which can fail to be true, as opposed to knowledge, which cannot. Allowing beliefs to be false in a "belief is probability 1" setting means that an agent can assign probability 1 to F even though F is false in the actual world. In such a circumstance, the agent must assign probability 0 to the negation $\neg F$. But then if we wish for the agent to statically revise her beliefs upon learning that her belief in F is mistaken, then what we seek is for the agent to revise on the event $\neg F$ that she assigns probability 0. Since we wish for this revision to proceed by way of conditionalization as per the fourth design principle, it follows that we seek a probabilistic conditionalization on an event of probability 0. This is forbidden in classical probability theory. However, it is possible in the Popper–Rényi–de Finetti theory of conditional probabilities (Popper 2002 [1935]; van Fraassen 1976, 1995; Réyni 1964, 1955; Halpern 2001), which takes conditional probability as its basic notion and provides an axiomatization of this notion. Connections between the Popper–Rényi–de Finetti theory and Belief Revision have been studied by van Fraassen (1995), Boutilier (1995), Halpern (2003), and Arló-Costa and Parikh (2005). In Section 1, we follow the presentation, definitions, and results of Baltag and Smets (2008a) in extending this work over finite spaces to DEL, thereby implementing the fifth and final design principle.

The axiomatic theories of conditional probability proposed by Popper (2002 [1935]), Rényi (1964), and van Fraassen (1976) and the "conditional lexicographic probability spaces" of Game Theory are all equivalent in the case of finite spaces.

Finite conditional probability space. A *finite conditional probability space* is a pair (W, μ) consisting of a nonempty finite set W of objects (called "worlds") and a function $\mu : \wp(W) \times \wp(W) \to [0,1]$ that maps any two sets of worlds $S, T \subseteq W$ to a real number $\mu(S \mid T)$ in the unit interval $[0,1]$ and that satisfies the axioms that make it a *Popper function* on W:

- $\mu(S \mid S) = 1$,
- $\mu(S \cup T \mid U) = \mu(S \mid U) + \mu(T \mid U)$ if $S \cap T = \emptyset$ and $U \neq \emptyset$, and
- $\mu(S \cap T \mid U) = \mu(S \mid T \cap U) \cdot \mu(T \mid U)$.

For convenience, we let $\mu(S)$ abbreviate $\mu(S \mid W)$.

Intuitively, $\mu(S \mid T)$ denotes the probability of S conditional on T. With our abbreviation $\mu(S) = \mu(S \mid W)$, the expression $\mu(S)$ denotes the unconditional probability of S. The unary unconditional probability function $\mu(- \mid W)$ is a classical probability function over our finite space W (i.e., the function satisfies Kolmogorov's axioms for finite spaces). Further, it can be shown that the classical conditionalization of this function coincides with the two-place conditional $\mu(- \mid -)$; that is,

$$\frac{\mu(S \cap T)}{\mu(T)} = \mu(S \mid T) \quad \text{if } \mu(T) \neq 0,$$

where the left side of the equality is the standard definition of the probability of S conditional on T in terms of the one-place classical probability function $\mu(-) := \mu(- \mid W)$ and the right side of the equality is just the value of the two-place Popper function $\mu(- \mid -)$. Accordingly, a Popper function is an extension of the classical notion of probability and conditional probability. Further, this extension does not prohibit conditionalization on events T of probability 0; that is, the Popper function μ assigns a real-number value $\mu(S \mid T) \in [0,1]$ even in the case that $\mu(T) = 0$. For example, taking $S = T = \emptyset$ and $U = W$ in the second Popper axiom, it follows that $\mu(\emptyset) = \mu(\emptyset \mid W) = 0$; however, setting $T = U = \emptyset$ in the third Popper axiom and observing that we have $\mu(\emptyset \mid \emptyset) = 1$ by the first Popper axiom, we have $\mu(S \mid \emptyset) = 1$. So it is possible to conditionalize on the probability-0 event $\emptyset$.

The Popper function μ of a finite conditional probability space is completely determined by the "relative priority" it assigns between pairs of worlds; that is, if we define the "relative priority of w over v" to be the value $(w, v)_\mu \in [0,1]$ obtained by setting

$$(w, v)_\mu := \mu(\{w\} \mid \{w, v\}),$$

then it follows that

$$\mu(S \mid T) = \sum_{w \in S \cap T} \left(\sum_{t \in T} \frac{(t,w)_\mu}{(w,t)_\mu} \right)^{-1} \quad \text{if } T \neq \emptyset,$$

where the usual conventions are used: $\sum_{x \in \emptyset} S_x = 0$, $\frac{1}{0} = \infty$, $\frac{1}{\infty} = 0$, $\infty + \infty = \infty$, and $x + \infty = \infty + x = \infty$ for each real number x. This reduction to "relative priority" of pairs suggests an alternative characterization of finite conditional probability spaces that, as we will see, will be useful for presenting a DEL-style account of probabilistic belief change.

Priority Space (Baltag and Smets 2008a). A *priority space* is a pair

$$(W, (-,-))$$

consisting of a nonempty finite set W of objects (called "worlds") and a function $(-,-) : W \times W \to [0,1]$ that maps any two worlds w and v to a real number (w,v) in the interval [0,1] and that satisfies the axioms that make it a *priority function* on W:

- $(w,w) = 1$ for all $w \in W$;
- $(w,v) = 1 - (v,w)$ for all $w, v \in W$ with $w \neq v$; and
- for all $u, v, w \in W$ such that $w \neq v$ and $(w,u) \cdot (u,v) + (v,u) \cdot (u,w) \neq 0$, we have

$$(w,v) = \frac{(w,u) \cdot (u,v)}{(w,u) \cdot (u,v) + (v,u) \cdot (u,w)}.$$

Priority Space Theorem (Baltag and Smets 2008a). There is a one-to-one correspondence mapping each finite conditional probability space (W,μ) to a priority space $(W,(-,-))$ such that $(w,v)_\mu = (w,v)$. In particular, given a Popper function μ on W, defining $(w,v) := (w,v)_\mu$ gives a priority function on W; and given a priority function $(-,-)$ on W, defining $\mu(S,T)$ by

$$\mu(S \mid T) := \begin{cases} \sum_{w \in S \cap T} \left(\sum_{t \in T} \frac{(t,w)_\mu}{(w,t)_\mu} \right)^{-1} & \text{if } T \neq \emptyset \\ 1 & \text{if } T = \emptyset \end{cases}$$

gives a Popper function on W.

We will use Popper functions (or their equivalent priority functions) to define and picture the models and action models for a DEL-style theory of probabilistic belief change. For each agent a, a Popper function μ_a will describe what the agent believes, conditionally or unconditionally. Just like in plausibility models, connected components of sets of worlds will be used to specify informational consistency and define knowledge. In particular, we will have the following:

- to say that "agent a knows F at world w" means that F is true at all worlds $x \in \mathsf{cc}_a(w)$ that are informationally equivalent with w according to agent a; and
- to say that "agent a believes F conditional on G at world w" means that a assigns probability 1 to the set of F-worlds conditional on the set of G-worlds that are informationally equivalent with w, meaning

$$\mu_a([\![F]\!] \mid [\![G]\!] \cap \mathsf{cc}_a(w)) = 1.$$

- to say that "agent a believes F (unconditionally) at world w" means that agent a believes F conditional on the propositional constant $\top$ for truth at world w.

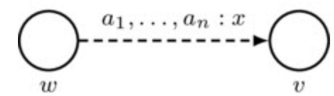
In the definition of conditional belief, we restrict the conditionalization to those worlds that are informationally equivalent with w to reflect the fact that we link knowledge with truth in all informationally equivalent worlds. In particular, if the actual world is w, then the agent knows that any world that is not informationally equivalent with w cannot be the actual world, and therefore she can rule out all such worlds. Accordingly, her beliefs must be consistent with what she knows, which requires conditionalization on subsets of worlds that are all informationally equivalent with the actual world.

> **Conditional probability model.** Given a nonempty set $\mathcal{P}$ of propositional letters and a finite nonempty set $\mathcal{A}$ of agents, a *conditional probability model* is a structure
>
> $$M = (W, \mu, \sim, V)$$
>
> consisting of a finite nonempty set W of *worlds* identifying the possible states of affairs that might obtain, a function μ that assigns to each agent a a Popper function μ_a on W, a function cc that assigns to each agent a an equivalence relation $\sim_a$ on W, and a *propositional valuation* V mapping each propositional letter to the set of worlds at which that letter is true. For worlds $w, v \in W$ and each agent $a \in \mathcal{A}$, we define the following:
>
> - $\mathsf{cc}_a(w) := \{v \in W \mid v \sim_a w\}$ is the set of worlds that are *informationally equivalent* to w according to agent a, and
> - $(w,v)_a := (w,v)_{\mu_a}$ is the relative priority of w over v according to agent a.
>
> A *pointed conditional probability model*, sometimes called a *scenario* or a *situation*, is a pair (M, w) consisting of a conditional probability model M and a world w (called the *point*) that designates the state of affairs that we (the formal modelers) currently assume to be actual.

We draw conditional probability models like plausibility models except that arrows are labeled differently. The labeling describes relative priorities agents assign between the worlds at the source and destination of the arrow. We require that arrows are labeled so as to respect the axioms of

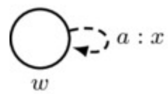
priority functions. This places restrictions on arrow labelings and leads to certain simplifications in our pictures.

- *Labeling Constraint*: arrow labels consist of nonempty finite sequences of agents, a colon, and a real number x in the unit interval $[0,1]$. In a picture:

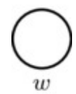
  $$\text{must satisfy} \begin{cases} n \in \mathbb{Z}^+ \\ a_1, \ldots, a_n \in \mathcal{A} \\ x \in [0,1] \end{cases}$$

  Intuitively, this arrow indicates that for each agent a_i appearing in the nonempty finite list $a_1, \ldots, a_n$ of agents, we have $(v,w)_{a_i} = x$, which says that agent a_i prioritizes the destination world v over the source world w with relative priority x. Note that conditional plausibility models can be distinguished from plausibility models by the addition of the colon symbol (":") and the real number $x \in [0,1]$ in the arrow label.
- *Labeling Totality and Uniqueness*: for each agent a and any two worlds w and v (including the possibility that $w = v$), there is a unique arrow from w to v containing a in its label, and there is a unique arrow from v to w containing a in its label. Note that not all arrows need be drawn, so long as implicit (i.e., undrawn) arrows can be uniquely inferred from the arrows that are drawn using the rules for drawing implicit arrows. These rules are presented as the next three bulleted items; there is one rule for each of the three axioms of priority spaces.
- *First Axiom*: arrows must respect the first axiom of priority spaces; that is, $(w,w)_a = 1$. In a picture:

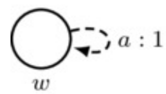


must satisfy $x = 1$

Also, for convenience, we adopt the following rule for simplifying drawings: for each agent a, there is an implicit reflexive arrow labeled by a and looping from each world w back to w itself such that the first axiom is satisfied.



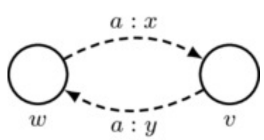
indicates



- *Second Axiom*: arrows must respect the second axiom of priority spaces; that is,

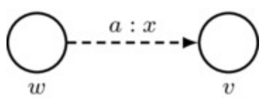
$$(w, v)_a = 1 - (v, w)_a \quad \text{if } w \neq v.$$

In a picture:

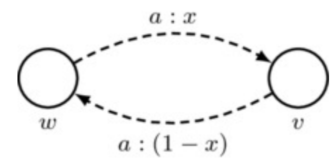


must satisfy $y = 1 - x$ if $w \neq v$

Also, for convenience, we adopt the following rule for simplifying drawings: for each agent a, each non-reflexive arrow that is labeled by a and the number x and is drawn in one direction gives rise to an implicit arrow that is labeled by a and the number $1 - x$ and is drawn in the opposite direction.



indicates

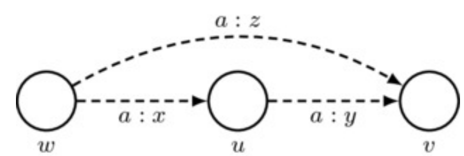


if $w \neq v$

- *Third Axiom*: arrows must respect the third axiom of priority spaces; that is, given $u, v, w \in W$ such that $w \neq v$ and $(w,u)_a \cdot (u,v)_a + (v,u)_a \cdot (u,w)_a \neq 0$, we have

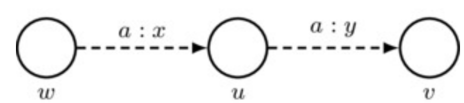
$$\begin{aligned}(w, v)_a &= \frac{(w,u)_a \cdot (u,v)_a}{(w,u)_a \cdot (u,v)_a + (v,u)_a \cdot (u,w)_a} \\ &= \frac{(w,u)_a \cdot (u,v)_a}{(w,u)_a \cdot (u,v)_a + (1 - (u,v)_a) \cdot (1 - (w,u)_a)},\end{aligned}$$

where the second equality follows by the first axiom of priority spaces. In a picture: if $w \neq v$ and $x \cdot y + (1 - y) \cdot (1 - x) \neq 0$, then

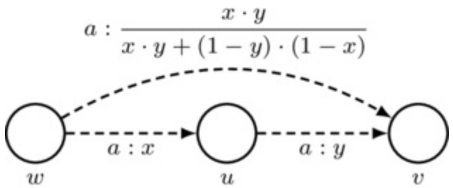


$$\text{must satisfy } z = \frac{x \cdot y}{x \cdot y + (1 - y) \cdot (1 - x)}$$

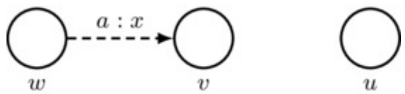
Also, for convenience, we adopt the following rule for simplifying drawings: for each agent a, every sequence of two a-labeled arrows that has different starting and ending worlds gives rise to an implicit third a-labeled arrow whose number is such that the third axiom of priority spaces is satisfied. Thus, if $w \neq v$ and $x \cdot y + (1 - y) \cdot (1 - x) \neq 0$, then



indicates



- *Informational Equivalence*: we write $w \sim_a v$ (equivalently that $w \in \mathsf{cc}_a(v)$ or that $v \in \mathsf{cc}_a(w)$) to mean that, if we ignore all arrowheads, then there is a sequence of zero or more dashed lines that create a path between w and v. So, for example, the picture



  indicates that $w \sim_a w$, $w \sim_a v$, $v \sim_a v$, $u \sim_a u$, $w \nsim_a u$, and $v \nsim_a u$ (along with each of the expressions $x \sim_a y$ and $x' \nsim_a y'$ appearing in this list with x and y exchanged).

The above items fully specify the rules and simplifications we may use to draw conditional probability models. We use these models to interpret the language (CPL) of Conditional Probability Logic.

$$F ::= p \mid F \wedge F \mid \neg F \mid [a : x]F \mid K_a F \qquad \text{(CPL)}$$
$$p \in \mathcal{P},\ a \in \mathcal{A},\ x \in [0,1]$$

The satisfaction relation $\vDash$ between pointed conditional probability models and formulas of (CPL) is defined as follows.

- $[\![F]\!] := \{w \in W \mid M, w \vDash F\}$.
- $M, w \vDash p$ holds if and only if $w \in V(p)$.
- $M, w \vDash F \wedge G$ holds if and only if both $M, w \vDash F$ and $M, w \vDash G$.
- $M, w \vDash \neg F$ holds if and only if $M, w \nvDash F$.
- $M, w \vDash [a : x]F$ holds if and only if $\mu_a([\![F]\!] \mid S \cap \mathsf{cc}_a(w)) \geq x$ for all $S \subseteq W$ such that $w \in S$.
- $M, w \vDash K_a F$ holds if and only if $M, v \vDash F$ for each $v \in \mathsf{cc}_a(w)$.

The formula $[a : x]F$ says, “agent a safely believes F with degree at least x”. This means that, conditional on any informationally consistent and truthful event, agent a assigns a probability no less than x to the statement that a is true. Note that this notion of “safe degree of belief” is different than the positive-integer notion $B_a^n F$ of degree-n belief introduced in the main article. Further, we observe that a variant of our notation $[a : x]F$ is to instead write $\Box_a^x F$.

For convenience, we let

$$\Box_a F \quad \text{abbreviate} \quad [a : 1]F,$$

so that the defeasible knowledge (or “safe belief”) formula $\Box_a F$ denotes degree-1 safe belief. With this abbreviation in hand, the language $(K\Box)$ can be interpreted over conditional probability models. This leads to the following result.

Probability-Plausibility Theorem (Baltag and Smets 2008a).

- Every conditional probability model $M = (W, \mu, \sim, V)$ gives rise to a plausibility model $M^- = (W, \geq, V)$ having the same worlds and valuation and satisfying the same $(K\Box)$-formulas: for each $w \in W$ and $F \in (K\Box)$, we have

$$M^-, w \vDash F \quad \text{iff} \quad M, w \vDash F.$$

  To obtain M^-, set $w \leq_a v$ if and only if $(w, v)_a \neq 0$.
- Every plausibility model $M = (W, \geq, V)$ gives rise to a conditional plausibility model $M^+ = (W, \mu, \sim, V)$ having the same worlds and valuation and satisfying the same $(K\Box)$-formulas: for each $w \in W$ and $F \in (K\Box)$, we have

$$M^+, w \models F \quad \text{iff} \quad M, w \models F.$$

To obtain M^+, use the Priority Space Theorem to define μ_a according to the following:

$$(w,v)_a := \begin{cases} 1 & \text{if } w <_a v \text{ or } w = v, \\ 0 & \text{if } w >_a v, \\ \frac{1}{2} & \text{otherwise.} \end{cases}$$

A consequence of the Probability-Plausibility Theorem is that the given model and the constructed model satisfy the *same knowledge and belief formulas*. This provides a natural connection between the quantitative static belief change setting of conditional probability models and the qualitative statice is statice correct? belief change setting of plausibility models. It also yields the following result.

> $\mathsf{K}\Box$ **Soundness and Completeness (Baltag and Smets 2008a).** The theory $\mathsf{K}\Box$ is sound and complete with respect to the collection C_* of pointed conditional probability models. That is, for each $(K\Box)$-formula F, we have $\mathsf{K}\Box \vdash F$ if and only if $C_* \models F$.

It has not yet been shown whether there is a complete axiomatization of the validities of the full language (CPL). It may also be of interest to determine whether there is such an axiomatization for the restricted language obtained by requiring all degrees of belief to be rational numbers in the unit interval $[0,1]$.

Baltag and Smets (2008a) also developed a probabilistic version of action models. These have a structure much like that of conditional probability models.

> **Conditional probability action model.** Given a set of formulas $\mathcal{L}$ and a finite nonempty set $\mathcal{A}$ of agents, a *conditional probability action model* is a structure
>
> $$A = (E, \mu, \mathsf{pre})$$
>
> consisting of a nonempty finite set E of the possible communicative *events* that might occur, a function μ that assigns to each agent a a Popper function μ_a on E, and a function pre that assigns to each event e in E a formula $\mathsf{pre}(e) \in \mathcal{L}$ called the *precondition* of e. A *pointed conditional probability action model*, sometimes also called an *action*, is a pair (A,e) consisting of a plausibility action model A and an event e in A that is called the *point*. Conditional probability action models are drawn like conditional probability models except that events are pictured as rectangles instead of as circles. The point of a conditional probability action model is indicated by a double rectangle.

We define language (CPAL) of Conditional Probability Action Logic along with the set CPAM_* of pointed conditional probability action models having preconditions in the language (CPAL) according to the following recursive grammar:

$$F ::= p \mid F \wedge F \mid \neg F \mid [a:x]F \mid K_a F \mid [A,e]F \qquad \text{(CPAL)}$$
$$p \in \mathcal{P},\ a \in \mathcal{A},\ x \in [0,1],\ (A,e) \in \mathsf{CPAM}_*$$

The satisfaction relation $\models$ between pointed conditional probability models and formulas of (CPAL) is an extension of the relation between $\models$ between pointed conditional probability models and formulas of (CPL) obtained by adding the following:

- $M, w \models [A,e]G$ holds if and only if $M, w \not\models \mathsf{pre}(e)$ or $M[A], (w,e) \models G$, where the model

  $$M[A] = (W[A], \mu[A], V[A])$$

  is the result of the *probabilistic product update* operation $M \mapsto M[A]$

defined by:

- $M[A] := \{(v,g) \in W \times E \mid M, v \models \mathsf{pre}(g)\}$ — pair worlds with events whose preconditions they satisfy;
- for each agent $a \in \mathcal{A}$, use the Priority Space Theorem to define $\mu[A]_a$ according to the following:

$$((v_1,g_1),(v_2,g_2))_a := \lim_{x \to (v_1,v_2)_a} \frac{x \cdot (g_1,g_2)_a}{x \cdot (g_1,g_2)_a + (1-x) \cdot (1-(g_1,g_2)_a)},$$

  where the limit is taken over all x's such that the denominator is nonzero; and
- $(v,g) \in V[A](p)$ if and only if $M, v \models p$ — make the truth of p at (v,g) like that at v.

The definition of the Popper function $\mu[A]_a$ resulting after the probabilistic product update by A may be understood by way of a three-part case analysis.

- In case $(g_1,g_2)_a = 0$, we have $((v_1,g_1),(v_2,g_2))_a = 0$.
- In case $(g_1,g_2)_a = 1$ and $g_1 \neq g_2$, we have $((v_1,g_1),(v_2,g_2))_a = 1$.
- In all other cases, we have

$$((v_1,g_1),(v_2,g_2))_a = \frac{(v_1,v_2)_a \cdot (g_1,g_2)_a}{(v_1,v_2)_a \cdot (g_1,g_2)_a + (1-(v_1,v_2)_a) \cdot (1-(g_1,g_2)_a)}.$$

The first two cases tell us the following: if g_1 and g_2 are distinct events and the agent believes conditional on one of g_1 or g_2 occurring that it is event g_i that occurred for some $i \in \{1,2\}$, then she will also believe conditional on one of (v_1,g_1) or (v_2,g_2) being the actual world that it is world (v_i,g_i) that is in fact actual. This is just the *action-priority update* from the language (APUL). This rule tells us that the incoming information from the conditional probability action model overrides information from the given conditional probability model whenever these two information sources are contradictory. This prioritization agrees with the basic assumption in Belief Revision theory that "new" incoming information (from the conditional probability action model) is to be prioritized over "old" information (from the conditional probability model) in case of contradiction. The third case above tells us that, so long as the "new" information does not contradict the "old" information, these two sources of information are to be considered as probabilistically independent, so the resulting relative priority of the newly created world (v_1,g_1) over the newly created world (v_2,g_2) is given by a product of the given relative priorities $(v_1,v_2)_a$ and $(g_1,g_2)_a$. The denominator on the right side of the equality of the third case ensures that this product is normalized over the subspace where the initial worlds to be considered are just v_1 and v_2 and the initial events to be considered are just g_1 and g_2.

It is an open question as to whether there is a complete axiomatization of the validities of (CPAL).

We refer the reader to Baltag and Smets (2008a) for further details, including the definition of a number of interesting conditional probability action models and the study of the game-theoretic notion of "common belief of rationality".

2. Probabilistic Dynamic Epistemic Logic

We now look at another approach to reasoning about probabilistic Dynamic Epistemic Logic due to van Benthem, Gerbrandy, and Kooi (2009b). This account begins with a different probabilistic semantics based on a different kind of model: the epistemic probability model.

Epistemic probability model. Given a nonempty set $\mathcal{P}$ of propositional letters and a finite nonempty set $\mathcal{A}$ of agents, an *epistemic probability model* is a structure

$$M = (W, \sim, P, V)$$

consisting of a finite nonempty set W of *worlds* identifying the possible states of affairs that might obtain, a function $\sim$ that assigns to each agent a an equivalence relation $\sim_a$ on W, a function P that assigns to each agent $a \in \mathcal{A}$ and world $w \in W$ a probability assignment $P_a(w) : W \to [0,1]$ to each world in W, and a *propositional valuation* $V : \mathcal{P} \to \wp(W)$ mapping each propositional letter to the set of worlds at which that letter is true. A *pointed epistemic probability model* is a pair (M, w) consisting of an epistemic probability model M and a world w (called the *point*) that designates the state of affairs that we (the formal modelers) currently assume to be actual.

The relation $\sim_a$ characterizes agent a's *knowledge*: if the actual wold is w, then "agent a knows F" is said to hold just in case F is true at all worlds $v \sim_a w$. This leads to the familiar $\mathsf{S5}$ notion of knowledge. The probability function describes agent a's subjective probabilistic belief as to which is the actual state of affairs: $P_a(w)(v)$ is the is the (subjective) probability agent a assigns to the event "the actual world is v" whenever the actual world is in fact w.

To reason about probability, authors such as Fagin, Halpern, and Megiddo (1990), Halpern (2003), Heifitz and Mongin (2001), and others have defined various formal languages that express the summed probability of a number of formulas is comparable in some way to (e.g., no greater than) a given number (usually rational, though real numbers may be allowed as well). Van Benthem, Gerbrandy, and Kooi (2009b) consider a simplified variant: the language (EPL) of *Epistemic Probabilistic Logic*. This language is defined according to the following grammar:

$$F ::= p \mid F \wedge F \mid \neg F \mid K_a F \mid P_a(F) = k \qquad \text{(EPL)}$$
$$p \in \mathcal{P},\ a \in \mathcal{A},\ k \in \mathbb{Q}$$

The formula $K_a F$ is read, "agent a knows F". The formula $P_a(F) = k$ is read, "agent a assigns probability k to F". The satisfaction relation $\models$ between pointed epistemic probability models and (EPL)-formulas is defined as follows.

- $[\![F]\!] := \{w \in W \mid M, w \models F\}$.
- $M, w \models p$ holds if and only if $w \in V(p)$.
- $M, w \models F \wedge G$ holds if and only if both $M, w \models F$ and $M, w \models G$.
- $M, w \models \neg F$ holds if and only if $M, w \not\models F$.
- $M, w \models K_a F$ holds if and only if $M, v \models F$ for each $v \sim_a w$.
- $M, w \models P_a(F) = k$ holds if and only if $k = \sum_{v \in [\![F]\!]} P_a(w)(v)$.

We note that epistemic probability models do not impose restrictions on the assignments of probability functions to states. However, in many applications, certain restrictions make sense, and one may wish to study the collection of (EPL)-validities that hold with respect to the given restricted class of epistemic probability models. As an example, many epistemic applications might wish to impose the property that

$$w \sim_a v \text{ implies } P_a(w)(v) = P_a(v)(w).$$

The class of pointed epistemic probability models satisfying this property validates the (EPL)-formula

$$(P_a(F) = k) \to K_a(P_a(F) = k),$$

which says that the agent knows the subjective probabilities that she assigns.

In examining DEL-style dynamics for epistemic probability models, van Benthem et al. (2009a) consider three sources of probabilistic information:

- *Prior probabilities*: these represent the agents' beliefs before receiving new information. Prior probabilities are just the agents' probability functions P_a coming from the epistemic probability model that occurs before the informational update.
- *Occurrence probabilities*: these represent objective frequencies about the process that produces the new information. Occurrence probabilities take the form of an assignment from information conveyed by the update (in the form of one of a finite number of logically inconsistent (EPL)-formulas) to probability functions over the finite set of possible events that might have brought about the update.
- *Observation probabilities*: these represent the agents' uncertainty as to the actual event that brought about the update. Observation probabilities take the form of an assignment from a finite set of events to probability functions over the set of events, indicating the agents' intrinsic uncertainty, independent of the particular formula-specified information the update conveys, as to which event is the actual one.

The action model-like structures defined by van Benthem, Gerbrandy, and Kooi (2009b) that realize informational updates with these three sources of probability are called *probabilistic update models*.

Probabilistic update model. Given a finite nonempty set $\mathcal{A}$ of agents, a *probabilistic update model* is a structure

$$A = (E, \sim, \Phi, \mathsf{pre}, P)$$

consisting of

- a nonempty finite set E of the possible communicative *events* that might occur,
- a function $\sim$ that assigns to each agent $a \in \mathcal{A}$ an equivalence relation $\sim_a$ on W,
- a finite nonempty set Φ of *precondition* formulas (in some language) that specify the possible information the update may convey,
- a function pre that assigns to each formula F a probability function $\mathsf{pre}(F, -) : E \to [0, 1]$ over E (for the occurrence probability), and
- a function P that assigns to each agent $a \in \mathcal{A}$ event $e \in E$ a probability function $P_a(w) : W \to [0, 1]$ over W (for the observation probability).

A *pointed probabilistic update model*, sometimes also called an *update*, is a pair (A, e) consisting of a probabilistic action model A and an event e in A that is called the *point*.

The language of *Epistemic Probabilistic Logic with Updates*, which we herein dub (EPL+U), along with the set EUM_* of pointed epistemic update models with preconditions in the language (EPL+U) according to the following recursive grammar:

$$F ::= p \mid F \wedge F \mid \neg F \mid K_a F \mid P_a(F) = k \mid [A, e]F \qquad \text{(EPL+U)}$$
$$p \in \mathcal{P},\ a \in \mathcal{A},\ k \in \mathbb{Q},\ (A, e) \in \mathsf{EUM}_*$$

The binary truth relation $\models$ between pointed epistemic probability models and (EPL+U)-formulas is obtained by extending the relation $\models$ for (EPL)-formulas by adding the following clause:

- $M, w \models [A, e]F$ holds if and only if for all $G \in \Phi$, we have $M, w \not\models G$ or $M[A]_b, (w, e) \models F$, where the model

$$M[A]_b = (W[A]_b, \sim[A]_b, P[A]_b, V[A]_b)$$

is the result of the van Benthem, Gerbrandy, and Kooi (2009b) probabilistic product update operation $M \mapsto M[A]_b$ defined by:

- For each $(v,g) \in W \times E$, we define the following quantity, which is just the occurrence probability of event g at world v, by setting

$$\mathsf{pre}(v,g) := \frac{\sum_{G \in \Phi} \{\mathsf{pre}(G,g) \mid v \in [\![G]\!]\}}{|\{G \in \Phi \mid v \in [\![G]\!]\}|}$$

  if the denominator (consisting of the number of preconditions satisfied by v) is nonzero and otherwise, if the denominator is zero, define $\mathsf{pre}(v,g) := 0$;
- $W[A]_b := \{(v,g) \in W \times E \mid \mathsf{pre}(v,g) > 0\}$ — pair worlds with events that have positive occurrence probability;
- $(w_1,e_1)(\sim[A]_b)_a(w_2,e_2)$ if and only if $w_1 \sim_a w_2$ in M and $e_1 \sim_a e_2$ in A — epistemic uncertainty is defined componentwise;
- $P[A]_B((w_1,e_1))((w_2,e_2))$ is defined to be

$$\frac{P_a(w_1)(w_2) \cdot \mathsf{pre}(w_2,e_2) \cdot P_a(e_1)(e_2)}{\sum_{(v,g) \in W \times E} P_a(w_1)(v) \cdot \mathsf{pre}(v,g) \cdot P_a(e_1)(g)}$$

  if the denominator is nonzero and defined to be 0 otherwise; and
- $(v,g) \in V[A]_b(p)$ if and only if $v \in V(p)$ — make the valuation of p at (v,g) as it was at v.

In van Benthem, Gerbrandy, and Kooi (2009b), it is assumed that preconditions are always mutually inconsistent; that is, no two distinct precondition formulas occurring in Φ are true at the same world. This assumption simplifies the definition above: if there is a unique precondition $G \in \Phi$ that is true at a world v, then $\mathsf{pre}(v,g)$ reduces to the value $\mathsf{pre}(G,g)$; otherwise, if there is no such unique precondition, then $\mathsf{pre}(v,g)$ is 0.

Van Benthem, Gerbrandy, and Kooi (2009b) show by way of a Reduction Theorem that the validities of a generalization of (EPL+U) obtained by considering linear inequalities

$$a_1 \cdot P_a(F_1) + \cdots + a_n \cdot P_a(F_n) \geq k$$

are axiomatizable. We refer the reader to van Benthem, Gerbrandy, and Kooi (2009b) for details. Also, for more general information about the consequences of expressing equalities with and without linear combinations, we point the reader to Section 3.2 of the *Stanford Encyclopedia of Philosophy* entry on Logic and Probability.

Appendix M: Preference dynamics

DEL-style model-changing operators have been applied by a number of researchers to the study of preferences, preference change, and related notions. At the semantic core of the various DEL approaches to reasoning about preference and preference change is some variant of the *preference model*.

> **Preference model.** Given a nonempty set $\mathcal{P}$ of propositional letters and a finite nonempty set $\mathcal{A}$ of agents, a *preference model* is a structure
>
> $$M = (W, \geq, V)$$
>
> consisting of a nonempty set W of *worlds* identifying the possible states of affairs that might obtain, a function $\geq$ that assigns to each agent a a reflexive and transitive binary relation$\geq_a$ on W, and a *propositional valuation* V mapping each propositional letter to the set of worlds at which that letter is true. A *pointed preference model*, sometimes called a *scenario* or a *situation*, is a pair (M,w) consisting of a preference model M and a world w (called the *point*) that

designates the state of affairs that we (the formal modelers) currently assume to be actual.

A preference model is just like a plausibility model except that the order $\geq_a$ need not satisfy the property of *Plausibility*; in particular, within each nonempty set of worlds, there need not always exist a nonempty subset consisting of the "most preferred" worlds. The expression $w \geq_a v$ is read, "agent a considers w to be at least as preferred as v".

Note that the convention here is that the "larger" worlds according to $\geq_a$ are more preferred; this is the opposite of the convention adopted for plausibility models. The converse preferential relation $\leq_a$, the strict preferential relations $\succ_a$ and $\prec_a$, and the equi-preferential relation $\simeq_a$ are defined in terms of $\geq_a$ in a manner analogous to our definitions of converse plausibility, strict plausibility, strict converse plausibility, and equi-plausibility in the definition of plausibility models.

Van Benthem et al. (2009) consider a single-agent version of a reflexive and irreflexive preference language we call (RIPL). This language consists of a universal modality $[\forall]$ ("true at all worlds in the model"), a reflexive preference modality $[\geq_a]$ ("true at the worlds non-strictly less preferred by a"), the converse reflexive preference modality $[\leq_a]$ ("true at the worlds non-strictly preferred by a"), an irreflexive preference modality $[\succ_a]$ ("true at the worlds strictly less preferred by a"), and the converse irreflexive preference modality $[\prec_a]$ ("true at the worlds strictly preferred by a").

$$\begin{aligned} F ::= {} & p \mid F \wedge F \mid \neg F \mid [\forall F] \mid \\ & [\geq_a]F \mid [\leq_a]F \mid [\succ_a]F \mid [\prec_a]F \\ & p \in \mathcal{P},\ a \in \mathcal{A} \end{aligned} \tag{RIPL}$$

We recall that the dual existential modality $[\exists]$ is defined by $[\exists]F := \neg[\forall]\neg F$. The dual modalities for the preference relation, which

are written between angled brackets, are defined similarly; for example, $\langle\geq_a\rangle F := \neg[\geq_a]\neg F$. The language (RIPL) is interpreted over pointed preference models as follows.

- $M, w \models p$ holds if and only if $w \in V(p)$.
- $M, w \models F \wedge G$ holds if and only if both $M, w \models F$ and $M, w \models G$.
- $M, w \models \neg F$ holds if and only if $M, w \not\models F$.
- $M, w \models [\forall]F$ holds if and only if $M, v \models F$ for each $v \in W$.
- $M, w \models [\geq_a]F$ holds if and only if $M, v \models F$ for each $v \geq_a w$.
- $M, w \models [\leq_a]F$ holds if and only if $M, v \models F$ for each $v \leq_a w$.
- $M, w \models [\succ_a]F$ holds if and only if $M, v \models F$ for each $v \succ_a w$.
- $M, w \models [\prec_a]F$ holds if and only if $M, v \models F$ for each $v \prec_a w$.

Van Benthem et al. (2009) show that (RIPL) allows us to express eight distinct notions of "agent a prefers G to F". A number of these notions arise in choosing best moves in a game-theoretic setting (van Benthem 2009).

Theorem (van Benthem et al. 2009). Let (M, w) be a pointed preference model.

1. Define $M, w \models F \leq_a^{\exists\exists} G$ to mean that $\exists x, \exists y \geq_a x : (M, x \models F$ and $M, y \models G)$.
   "There is an F-world that agent a non-strictly prefers to some G-world."
   The formula $[\exists](F \wedge \langle\leq_a\rangle G)$ is equivalent to $F \leq_a^{\exists\exists} G$.
2. Define $M, w \models F \leq_a^{\forall\exists} G$ to mean that $\forall x, \exists y \geq_a x : (M, x \models F$ implies $M, y \models G)$.
   "For every F-world, there is a G-world that agent a non-strictly prefers."
   The formula $[\forall](F \rightarrow \langle\leq_a\rangle G)$ is equivalent to $F \leq_a^{\forall\exists} G$.
3. Define $M, w \models F <_a^{\exists\exists} G$ to mean that $\exists x, \exists y \succ_a x : (M, x \models F$

and $M,y \vDash G$).

"There is an F-world that agent a strictly prefers to some G-world."

The formula $[\exists](F \wedge \langle \prec_a \rangle G)$ is equivalent to $F <_a^{\exists\exists} G$.

4. Define $M,w \vDash F <_a^{\forall\exists} G$ to mean that $\forall x, \exists y$: (if $M,x \vDash F$, then $M,y \vDash G$ and $y \succ_a x$).

   "For every F-world, there is a G-world that agent a strictly prefers."

   The formula $[\forall](F \to \langle \prec_a \rangle G)$ is equivalent to $F <_a^{\forall\exists} G$.
5. Define $M,w \vDash F <_a^{\forall\forall} G$ to mean that $\forall x, \forall y$: (if $M,x \vDash F$ and $M,y \vDash G$, then $x \prec_a y$).

   "Agent a strictly prefers every G-world to every F-world."

   If the preference ordering $\succeq_a$ is total (see Appendix C), then $[\forall](G \to [\preceq_a]\neg F)$ is equivalent to $F <_a^{\forall\forall} G$.
6. Define $M,w \vDash G >_a^{\exists\forall} F$ to mean that $\exists y, \forall x$: (if $M,x \vDash F$ and $M,y \vDash G$, then $x \prec_a y$).

   "There is a G-world that agent a strictly prefers to every F-world."

   If the preference ordering $\succeq_a$ is total, then $[\exists](G \wedge [\preceq_a]\neg F)$ is equivalent to $G >_a^{\exists\forall} F$.
7. Define $M,w \vDash F \leq_a^{\forall\forall} G$ to mean that $\forall x, \forall y$: (if $M,x \vDash F$ and $M,y \vDash G$, then $x \preceq_a y$).

   "Agent a non-strictly prefers every G-world to every F-world."

   If the preference ordering $\succeq_a$ is total, then $[\forall](G \to [\prec_a]\neg F)$ is equivalent to $F \leq_a^{\forall\forall} G$.
8. Define $M,w \vDash G \geq_a^{\exists\forall} F$ to mean that $\exists y, \forall x$: (if $M,x \vDash F$ and $M,y \vDash G$, then $x \preceq_a y$).

   "There is a G-world that agent a non-strictly prefers to every F-world."

   If the preference ordering $\succeq_a$ is total, then $[\exists](G \wedge [\prec_a]\neg F)$ is equivalent to $G \geq_a^{\exists\forall} F$.

To consider multi-agent preferences in conjunction with multi-agent knowledge, and to allow for DEL-style changes in preferences and knowledge, van Benthem and Liu (2007) study preference models to which epistemic relations R_a are added for each agent a.

> **Epistemic preference model.** Given a nonempty set $\mathcal{P}$ of propositional letters and a finite nonempty set $\mathcal{A}$ of agents, an *epistemic preference model* is a structure
>
> $$M = (W, \succeq, R, V)$$
>
> consisting of a nonempty set W of *worlds* identifying the possible states of affairs that might obtain, a function $\succeq$ that assigns to each agent $a \in \mathcal{A}$ a reflexive and transitive binary relation $\succeq_a$ on W, a function R that assigns a binary equivalence relation R_a on W to each agent $a \in \mathcal{A}$, and a *propositional valuation* V mapping each propositional letter to the set of worlds at which that letter is true. A *pointed epistemic preference model*, sometimes called a *scenario* or a *situation*, is a pair (M,w) consisting of an epistemic preference model M and a world w (called the *point*) that designates the state of affairs that we (the formal modelers) currently assume to be actual.

We note that van Benthem and Liu (2007) define epistemic preference models so that each R_a is an equivalence relation. This is because they wish to adopt the standard logic of knowledge (multi-agent $\mathsf{S5}$) and assign formulas $[a]F$ an epistemic reading ("agent a knows F"). This restriction that the R_a's be equivalence relations is not a technical necessity and could easily be varied if other normal modal logics for the modalities $[a]$ are desired.

Van Benthem and Liu (2007) define the single-agent dynamic epistemic preference language (DEPL) that makes use of a universal modality, reflexive preference formulas $[\preceq_a]F$ ("agent a prefers F") for each agent

a, a knowledge modality $[a]F$ ("agent a knows F") for each agent a, "link-cutting" public announcement formulas $[F!']G$ ("after the announcement of F, formula G is true"), and "preference upgrade" formulas $[\sharp F]G$ ("after eliminating preferences for $\neg F$-worlds over F-worlds, G is true"). Here we consider a simple multi-agent extension.

$$\begin{aligned} F ::= {} & p \mid F \wedge F \mid \neg F \mid [a]F \mid \\ & [\forall]F \mid [\preceq_a]F \mid [F!']F \mid [\sharp F]F \\ & p \in \mathcal{P},\ a \in \mathcal{A} \end{aligned} \qquad \text{(DEPL)}$$

Formulas of this language are interpreted at pointed epistemic preference models using the appropriate clauses from (RIPL) and from (ML) along with the following clauses:

- $M, w \models [a]F$ holds if and only if $M, v \models F$ for each v satisfying $wR_a v$.
- $M, w \models [F!']G$ holds if and only if we have that $M[F!'], w \models G$, where the model

  $$M[F!'] = (W[F!'], \geq[F!'], R[F!'], V[F!'])$$

  is defined by:
  - $W[F!'] := W$ — retain all worlds,
  - $x \geq [F!']_a y$ if and only if $x \geq_a y$ — leave preferences as before,
  - $xR[F!']_a y$ if and only if $xR_a y$ and we have $M, x \models F$ if and only if $M, y \models F$ — delete only those epistemic connections between F-worlds and $\neg F$-worlds, and
  - $v \in V[F!'](p)$ holds if and only if $v \in V(p)$ — leave the valuation the same at all worlds.
- $M, w \models [\sharp F]G$ holds if and only if we have that $M[\sharp F], w \models G$, where the model

  $$M[\sharp F] = (W[\sharp F], \geq[\sharp F], R[\sharp F], V[\sharp F])$$

  is defined by:
  - $W[\sharp F] := W$ — retain all worlds,
  - $x \geq [\sharp F]_a y$ if and only if $x \geq_a y$ and it is not the case that both $M, w \not\models F$ and $M, y \models F$ — delete only those preferences for $\neg F$-worlds over F-worlds,
  - $xR[\sharp F]_a y$ if and only if $xR_a y$ — leave epistemic relations as before, and
  - $v \in V[\sharp F](p)$ if and only if $v \in V(p)$ — leave the valuation the same at all worlds.

Van Benthem and Liu (2007) axiomatize the (DEPL)-validities and call the resulting logic Dynamic Epistemic Upgrade Logic DEUL.

The axiomatic theory DEUL.

- Axiom schemes and rules for classical propositional logic
- S5 axiom schemes and rules for $[a]$
- S4 axiom schemes and rules for $[\preceq_a]$
- Universal modality axioms:
  - S5 axiom schemes and rules for $[\forall]$
  - $[\forall]F \to [a]F$
  - $[\forall]F \to [\preceq_a]F$
- Link-cutting public announcements:
  - $[F!']p \leftrightarrow (F \to p)$ for letters $p \in \mathcal{P}$
    "After a false announcement, every letter holds—a contradiction. After a true announcement, letters retain their truth values."
  - $[F!'](G \wedge H) \leftrightarrow ([F!']G \wedge [F!']H)$
    "A conjunction is true after an announcement iff each conjunct is."
  - $[F!']\neg G \leftrightarrow (F \to \neg[F!']G)$
    "G is false after an announcement iff the announcement,

whenever truthful, does not make G true."

- $[F!'][a]G \leftrightarrow (F \to [a][F!']G)$
  "a knows G after an announcement iff the announcement, whenever truthful, is known by a to make G true."
- $[F!'][\leq_a]G \leftrightarrow (F \to [\leq_a][F!']G)$
  "a prefers G after an announcement iff a prefers that G is true after the announcement is truthfully made."
- $[F!'][\forall]G \leftrightarrow (F \to [\forall]([F!']G \wedge [\neg F!']G))$
  "G is true everywhere after an announcement of F iff whenever F is true, it is true everywhere that both the announcement of F and the announcement of its negation makes G true."
- Link-Cutting Announcement Necessitation Rule: from F, infer $[G!']F$.
  "A validity holds after any announcement."

- Preference upgrades:
  - $[\sharp F]p \leftrightarrow p$ for letters $p \in \mathcal{P}$
    "p holds after an upgrade iff p held before the upgrade."
  - $[\sharp F](G \wedge H) \leftrightarrow ([\sharp F]G \wedge [\sharp F]H)$
    "A conjunction is true after an upgrade iff each conjunct is."
  - $[\sharp F]\neg G \leftrightarrow \neg[\sharp F]G$
    "A negation is true after an upgrade iff it is false before the upgrade."
  - $[\sharp F][a]G \leftrightarrow [a][\sharp F]G$
    "a knows G after an upgrade iff the upgrade is known by a to make G true."
  - $[\sharp F][\leq_a]G \leftrightarrow ([\leq_a](F \to [\sharp F]G) \wedge (\neg F \to [\leq_a][\sharp F]G))$
    "a prefers G after an upgrade by F iff a prefers that 'the upgrade by F make G true at F-worlds,' and, in case F is false, a also prefers that the upgrade of F make G true."
  - $[\sharp F][\forall]G \leftrightarrow [\forall][\sharp F]G$

    "G is true everywhere after an upgrade of F iff it is true everywhere that the upgrade makes G true."
  - Preference Upgrade Necessitation Rule: from F, infer $[\sharp G]F$.
    "A validity holds after any upgrade."

DEUL **Soundness and Completeness (van Benthem and Liu 2007).** DEUL is sound and complete with respect to the collection $\mathcal{C}_*$ of pointed epistemic preference models. That is, for each (DEPL)-formula F, we have that DEUL $\vdash F$ if and only if $\mathcal{C}_* \models F$.

The work of van Benthem and Liu (2007) has been further developed by a number of authors. Liu (2008) looks at a quantitative version of preference and preference change closely related to earlier work on belief revision by Aucher (2003). Yamada (2007a,b, 2008) examines various deontic logics of action, command, and obligation. Van Eijck (2008) looks at a generalized Propositional Dynamic Logic-style preference logic that encompasses van Benthem and Liu (2007) and allows for the study of common knowledge (or belief) along with a DEL-style notion of conditional belief and belief change. Van Eijck and Sietsma (2010) further extend this work, examining applications to judgment aggregation; this has natural connections with coalition logic and social choice theory. Van Benthem, Girard, and Roy (2009c) examine a preference logic for von Wright's (1963) notion of *ceteris paribus* (in the distinct senses of "all things being normal" and of "all things being equal") along with a dynamic notion of "agenda change". A textbook on preferences and preference dynamics is Liu (2011).

Appendix N: Temporal aspects of Dynamic Epistemic Logic

Sack (2007, 2008, 2010) and Yap (2006, 2011) have suggested an extension of action model languages obtained by adding new temporal formulas $[Y]F$ assigned the reading "F was true yesterday (i.e., one time-step ago)". The suggestion is to use the so-called "yesterday" modal operator $[Y]$ to describe the passage of time explicitly in the language. For instance, if we wish to formalize our suggestion that time flows forward whenever actions occur, then we might wish to endorse a principle such as

$$p \to [A,e][Y]p,$$

which says, "if p is true, then, after the action occurs, we have that p was true yesterday". Intuitively, this makes sense: if p is true before the occurrence of the event, then it must be the case that p is true just after the event occurs. Formally endorsing this principle (i.e., developing a formal semantics and an axiomatics that respects the appropriate reduction axioms and in which the above formula is valid for each letter p) is tantamount to enforcing the requirement that every action has one time-step (or "clock-tick") in duration. That is, the passage of time and the occurrence of actions are directly linked: to ask what time it is amounts to asking how many actions have occurred since some fixed starting point. But there are other possible linkages one might wish to consider as well. Letting $[Y]^n F$ denote the formula F proceeded by n occurrences of the yesterday modal $[Y]$ (i.e., $[Y]^0 F := F$ and $[Y]^{n+1} F = [Y]^n [Y] F$), what follows are a few examples.

- Direct linkage: with the occurrence of each action, the clock ticks some fixed nonzero number of times n. We can think of n as the "duration" of the action. (Above we discussed the case $n = 1$.) Therefore, to ask what time it is under direct linkage amounts to asking for the value of n times the number of actions that have occurred since some fixed moment. In symbols, accepting direct linkage amounts to formally endorsing the principle

$$M, w \vDash p \to [A,e][Y]^n p.$$

- Proportional linkage: with the occurrence of each action, the clock ticks a number of times equal to a function f of the size $s(A,e)$ of the action. So the duration of an action depends in some way on its size. Therefore, to ask what time it is under proportional linkage amounts to asking for the sum of the durations of all actions that have occurred since some fixed moment. In symbols, accepting proportional linkage amounts to formally endorsing the principle

$$M, w \vDash p \to [A,e][Y]^{f(s(A,e))} p.$$

- Situational linkage: with the occurrence of each action, the clock ticks a number of times equal to a function g that depends on the action A as considered within the context of the current situation (M,w) in which action A occurs. So the duration of an action (and even the issue of whether an action has some nonzero duration) depends on the action and the situational context in which the action occurs. Therefore, to ask what time it is under situational linkage amounts to asking how many times "affirmative clock-tick conditions" have been met (as governed by the output of the clock tick-specifying function g). In symbols, accepting situational linkage amounts to formally endorsing the principle

$$M, w \vDash p \to [A,e][Y]^{g(M,w,A,e)} p.$$

Proportional linkage is a generalization of direct linkage (take f constant), and situational linkage is a generalization of proportional linkage (take $g(M,w,A,e)$ to be the value of f on the size $s(A,e)$).

Let us assume we have fixed one particular relationship between the flow of time and the occurrence of model-changing events. For example, following Sack (2008, 2010) and Yap (2006, 2011), let us suppose we pick direct linkage with $n = 1$: time flows one clock-tick per action. This fixes an "objective" relationship between time and the occurrence of actions. That is, according to the assumption of direct linkage with $n = 1$, the objective fact is that the time is given in terms of the number of actions that have occurred since some particular fixed starting point. A different linkage would lead to a logical framework with a different objective factual accounting of time. But note that the agents' "subjective experience" of the linkage between time and the occurrence of actions need not match the objective truth. For example, perhaps the subjective experience is *synchronous*, wherein the agents have no disagreement, mistakeness, or uncertainty with regard to the objective measure of time. Or perhaps it satisfies the property of *perfect recall*, which in one variation says that no agent forgets her past knowledge after the occurrence of an action. Other properties might be satisfied as well. This leads us to the following question: which combinations of objective linkage and subjective experience are logically compatible with a particular axiomatic theory of DEL?

Some work has been done on this question. Yap (2006, 2011), van Benthem, Gerbrandy, and Pacuit (2007), Sack (2008, 2010), Hoshi (2009), Hoshi and Yap (2009), van Benthem et al. (2009a), and Wang and Cao (2013), and looked at $n = 1$ direct linkage satisfying synchronicity, perfect recall, and other interesting subjective properties. See Hoshi (2010) for a survey of the traditional approach in DEL; see Wang and Cao (2013) for an alternative axiomatic approach more in the spirit of temporal logic. These lines of work generally follow the suggestion we outlined above: time flows one clock-tick per model-changing event, and the agents have common knowledge of this fact (though individual knowledge with respect to other atemporal issues may vary). So this is a synchronous, perfect recall-satisfying $n = 1$ direct linkage with a matching subjective experience of $n = 1$ direct linkage. Dégremont, Löwe, and Witzel (2011) introduce an $n = 1$ direct linkage in an asynchronous setting. In this approach, time flows just as before, one clock-tick per model-changing event. The difference is in the subjective experience: an agent knows a model-changing event has occurred if her knowledge or belief of something unrelated to time (or "atemporal") changes. For example, suppose p is true but agent a does not know whether this is so. After the public announcement of p, agent a's knowledge changes because she comes to learn p, a statement that not that does not contain a temporal modal $[Y]$ and that is therefore unrelated to time (i.e., atemporal). Therefore, the agent will know that the public announcement of p has occurred, and hence she will know that the clock ticked once as per this announcement. But then if the public announcement of p occurs again, agent a's atemporal knowledge will not change: she already knows p (since it was just announced) and therefore her knowledge about things unrelated to time does not change as per this second announcement of p. Therefore, while the clock will tick a second time as per the occurrence of the second announcement (and the assumption of $n = 1$ direct linkage), she will not know that it has indeed ticked because her atemporal knowledge and belief remains the same. This leads to an asynchronous state wherein she is confused as to the actual time, which is objectively two clock-ticks later but subjectively either one or two clock-ticks later. This gives the general idea of the approach due to Dégremont, Löwe, and Witzel (2011), an asynchronous $n = 1$ direct linkage with a subjective experience of situational linkage. Finally, we mention the approach of Renne, Sack, and Yap (2009, 2015), which introduces a generalized framework for reasoning about a variety of direct linkages and subjective experiences all within the same logic.

The Dynamic Epistemic Logic approach to reasoning about model change has natural connections with the *interpreted systems* approach of Fagin et al. (1995) and the *Epistemic Temporal Logic* approach of Parikh and Ramanujam (2003), though there are some important differences. In particular, in these non-DEL approaches, as in temporal logic in general, all temporal information is included from the start within the model. Intuitively, the model contains all information about the past, present, and future. Contrast this with DEL, where models represent the present and, in some cases, the past as well. But in either case, DEL theories developed to date leave the future "open" in the sense that future states of affairs are realized not by looking within a given model but instead by applying a model-changing action modality. It is such model-changing actions that produce "future" states of affairs. So in the "open future" approach of DEL, different future states of affairs are realized by way of different sequences of action model modalities. In "closed future" approaches (e.g., interpreted systems, Epistemic Temporal Logic, and temporal logic), future states of affairs are all together described in advance within a single model. See Hoshi (2010) and Dégremont, Löwe, and WWitzel (2011) for more details.

Another interesting area of work in DEL is to extend "closed future" theories with DEL-style model-changing operators, which introduces many technical and philosophical challenges (e.g., what do we mean by "future" if there are both formulas $[T]F$ that look at what is true at a static "tomorrow"-world found within the current model and formulas $[A,e]F$ that look at what is true after the occurrence of a dynamic, model-changing operation?). The work of Renne et al. (2009, 2015) presents an approach for updating Epistemic Temporal Logic-style models; however, this approach avoids complications by restricting future operators to the dynamic action model modalities $[A,e]$, effectively ignoring "static" futures that would otherwise only be accessible via a "tomorrow" modal operator $[T]$.