Information dynamics and uniform substitution [*]

Wesley H. Holliday (`wesholliday@berkeley.edu`)
Department of Philosophy, University of California, Berkeley
Logical Dynamics Lab, Center for the Study of Language and Information

Tomohiro Hoshi (`thoshi@stanford.edu`)
Educational Program for Gifted Youth, Stanford University
Logical Dynamics Lab, Center for the Study of Language and Information

Thomas F. Icard, III (`icard@stanford.edu`)
Department of Philosophy, Stanford University
Logical Dynamics Lab, Center for the Study of Language and Information

Abstract. The picture of information acquisition as the elimination of possibilities has proven fruitful in many domains, serving as a foundation for formal models in philosophy, linguistics, computer science, and economics. While the picture appears simple, its formalization in *dynamic epistemic logic* reveals subtleties: given a valid principle of information dynamics in the language of dynamic epistemic logic, substituting complex epistemic sentences for its atomic sentences may result in an invalid principle. In this article, we explore such failures of *uniform substitution.* First, we give epistemic examples inspired by Moore, Fitch, and Williamson. Second, we answer affirmatively a question posed by van Benthem: can we effectively decide when every substitution instance of a given dynamic epistemic principle is valid? In technical terms, we prove the decidability of this *schematic validity* problem for Public Announcement Logic (PAL and PAL-RC) over models for finitely many fully introspective agents, as well as models for infinitely many arbitrary agents. The proof of this result illuminates the reasons for the failure of uniform substitution.

Keywords: dynamic epistemic logic, public announcement logic, uniform substitution, schematic validity, substitution core, decidability

1. Introduction

The concept of an agent's *information state*, understood as the set of possibilities compatible with the agent's current information, has gained currency in philosophy, linguistics, computer science, and economics. So has the dynamic picture of information acquisition as the elimination of possibilities from the agent's information state. When enriched by relations between the information states of different agents, a space of possibilities can also represent an agent's uncertainty about what information other agents possess. It can even represent an agent's uncertainty about her own information state. In each case, a reduction of uncertainty comes with the elimination of certain possibilities.

[*] Published in *Synthese*, Vol. 190, 31-55, 2013. DOI 10.1007/s11229-013-0278-0.

Suppose agent i is an infallible source of information for agent j, in the following sense: i only sends j a message when it is true, and as soon as i does so, all possibilities incompatible with the truth of the message are eliminated for j. In this way, j learns from i. However, what j learns is not necessarily that the message from i is now true, even if it *is* still true. For example, we claim that there are ways of filling in the blanks below (uniformly within each line) such that:

- i tells j "____," but then j doesn't know (the truth) that ____;
- j initially knows ____, but then as a result of being told "____" by i, j doesn't know (the truth) that ____.

Later we will give concrete examples of these forms. Try filling in the blanks with descriptions of *j's current information* (or lack thereof).

Let us relate the problem of filling in the blanks to the logical notion of *uniform substitution*. For a given logical system, the set of valid formulas is *closed under uniform substitution* iff for any valid formula $\varphi(p)$ containing an atomic sentence p, the result $\varphi(\psi/p)$ of substituting an arbitrary formula ψ for all occurrences of p in φ ("filling in the blanks") is also a valid formula. Most traditional logical systems, including epistemic logics [24, 20, 33] for reasoning about the knowledge of agents, are closed under uniform substitution in this sense.

However, when we turn to *dynamic* epistemic logics [35, 22, 4, 9, 19, 8] for reasoning about the kind of information change that occurs when j acquires knowledge from i, we lose closure under uniform substitution. The reason is that for some *valid* principles $\varphi(p)$ of these logics, substituting an epistemic formula ψ about an agent's current information for an atomic sentence p results in an *invalid* substitution instance $\varphi(\psi/p)$.[1] This is the connection between information dynamics and uniform substitution that we will explore in the rest of this article.

We will begin by reviewing the basics of dynamic epistemic logic: Public Announcement Logic (PAL) [35, 22] in §1.1 and its extension with relativized common knowledge, PAL-RC [9], in §1.3. In §1.2 and

[1] Dynamic epistemic logics are not the only modal logics to have been proposed that are not closed under uniform substitution. Other examples include Buss's [11] modal logic of "pure provability," Åqvist's [1] two-dimensional modal logic (see [39]), Davies and Humberstone's [14] two-dimensional logic of "actually" and "fixedly," Carnap's [12] modal system for logical necessity (see [3, 38]), an epistemic-doxastic logic proposed by Halpern [23], and the full computation tree logic CTL* (see [37]). Among propositional logics, inquisitive logic [32, 13] is a non-uniform example, as is the combined classical-intuitionistic logic of del Cerro and Herzig [15]. In some cases, the substitution-closed set of validities—the *substitution core*—turns out to be another known system. For example, the substitution core of Carnap's system **C** is **S5** [38], and the substitution core of inquisitive logic is Medvedev Logic [13, §3.4].

§3, we use PAL to model concrete examples of the forms described by the bullet points above, showing how PAL is not closed under uniform substitution. Given that some valid dynamic epistemic principles have invalid substitution instances, the question naturally arises [7, 6, 9]:

> What are the valid dynamic epistemic principles (of PAL and PAL-RC) all of whose substitution instances are valid? Is the set of such "schematically valid" principles even *decidable*?

The decidability question is Question 1 in van Benthem's list of "Open Problems in Logical Dynamics" [7]. In §2, we answer this question affirmatively for the systems of PAL and PAL-RC over models for finitely many fully introspective agents (with transitive and Euclidean accessibility relations), as well as models for infinitely many agents with or without introspection (arbitrary relations). The proof of this result illuminates the reasons for the failure of uniform substitution. Elsewhere we give a complete axiomatization of the substitution-closed fragment with a system of Uniform Public Announcement Logic (UPAL) [29].

1.1. Dynamic Epistemic Logic

In this section, we review the simplest system of dynamic epistemic logic, designed to reason about information acquisition as the elimination of possibilities: Public Announcement Logic (PAL) [35, 22].

The language of PAL, $\mathcal{L}_{\mathsf{PAL}}$, extends that of multi-agent epistemic logic. For a set $\mathsf{At} = \{p, q, \dots\}$ of atomic sentences and a set $\mathsf{Agt} = \{i, j, \dots\}$ of agent symbols, $\mathcal{L}_{\mathsf{PAL}}$ is defined by the following grammar:

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid K_i\varphi \mid \langle\varphi\rangle\varphi,$$

where $p \in \mathsf{At}$ and $i \in \mathsf{Agt}$. The set of atomic sentences in φ is $\mathsf{At}(\varphi)$.

We take each atom $p, q, \dots$ of the language to stand for an "eternal" sentence in the sense of Quine [36, §40]. For example, 'Sacramento became the capital of California in 1854' is such an eternal sentence, whereas 'Sacramento is the capital of California' is not. Given a sentence φ, we read $K_i\varphi$ as the present tense ascription: *agent i knows* (*or has the information*) *that φ*. Hence the static part of our language, without the $\langle\varphi\rangle$ operators, consists of what could be called PKEA sentences for present knowledge, eternal atoms. We will say that such a PKEA sentence ψ can be true at one time and false at another time.[2] Given sentences φ and ψ, we read $\langle\varphi\rangle\psi$ as: *after all agents publicly receive the true information that φ, ψ*. For example, where p stands

[2] Those who do not like to speak this way about sentences should understand us as saying that ψ can be *truly uttered* at the one time but not at the other.

for the eternal sentence about the capital of California, we read $\langle p\rangle K_i p$ as follows: after all agents publicly receive the true information that Sacramento became the capital of California in 1854, agent i knows that Sacramento became the capital of California in 1854.

Models for PAL are the same as models for multi-agent epistemic logic. They are relational structures of the form

$$\mathcal{M} = \langle W, \{R_i \mid i \in \mathsf{Agt}\}, V\rangle,$$

where W is a set of possibilities, R_i is agent i's binary "epistemic accessibility" relation on W, and V is a valuation function mapping each atomic sentence p to a subset of W. For $w, v \in W$, we take wR_iv to mean that v is consistent with agent i's information in w.

We can now define agent i's *information state* in w as

$$R_i(w) = \{v \in W \mid wR_iv\},$$

the set of possibilities consistent with agent i's information in w, and say that i "knows" φ if and only if φ is true throughout this set.[3]

The truth clauses for the static part of the language are:

$$\begin{array}{lll}
\mathcal{M}, w \vDash p & \text{iff} & w \in V(p); \\
\mathcal{M}, w \vDash \neg\varphi & \text{iff} & \mathcal{M}, w \nvDash \varphi; \\
\mathcal{M}, w \vDash (\varphi \wedge \psi) & \text{iff} & \mathcal{M}, w \vDash \varphi \text{ and } \mathcal{M}, w \vDash \psi; \\
\mathcal{M}, w \vDash K_i\varphi & \text{iff} & \forall v \in R_i(w)\colon \mathcal{M}, v \vDash \varphi.
\end{array}$$

We denote the extension of φ in $\mathcal{M}$ by $[\![\varphi]\!]^{\mathcal{M}} = \{v \in W \mid \mathcal{M}, v \vDash \varphi\}$.

Given our informational interpretation of R_i, it is natural to assume that R_i is at least a *reflexive* relation, so an agent's "information" is *true* information ($K_i\varphi \to \varphi$). In many applications, it is also assumed that R_i is transitive and Euclidean, reflecting the idealization that the agent knows what information she has ($K_i\varphi \to K_iK_i\varphi$) and doesn't have ($\neg K_i\varphi \to K_i\neg K_i\varphi$). We call a model for such fully introspective agents with transitive and Euclidean relations a *quasi-partition* model. If all of the relations in a quasi-partition model are also reflexive—if they are equivalence relations—then we call it a *partition* model.

To define truth for the dynamic part of the language—with $\langle\varphi\rangle$ operators—we need the following crucial definition to formally capture the picture of information acquisition as the elimination of possibilities.

Definition 1 (Public Information Update) Given a model $\mathcal{M} = \langle W, \{R_i \mid i \in \mathsf{Agt}\}, V\rangle$, we obtain the updated model

$$\mathcal{M}_{|\varphi} = \langle W_{|\varphi}, \{R_{i_{|\varphi}} \mid i \in \mathsf{Agt}\}, V_{|\varphi}\rangle$$

[3] We do not pretend that this idealized model captures all the nuances of the notion of knowledge studied in epistemology. For discussion of epistemic logic and epistemology, see the references [27, 26].

by *eliminating from W all possibilities in which φ was false*:

- $W_{|\varphi} = [\![\varphi]\!]^{\mathcal{M}}$;
- for all $i \in \mathsf{Agt}$, $R_{i_{|\varphi}} = R_i \cap (W_{|\varphi} \times W_{|\varphi})$;
- for all $p \in \mathsf{At}$, $V_{|\varphi}(p) = V(p) \cap W_{|\varphi}$.

Using this definition, we can now state the truth clause that makes PAL a *dynamic* epistemic logic and that explains our gloss of $\langle\varphi\rangle\psi$ as "*after* all agents publicly receive the true information that φ, ψ":

$$\mathcal{M}, w \vDash \langle\varphi\rangle\psi \text{ iff } \mathcal{M}, w \vDash \varphi \text{ and } \mathcal{M}_{|\varphi}, w \vDash \psi.$$

In other words, the formula $\langle\varphi\rangle\psi$ is true at w just in case, first, φ is true at w in the *initial* model, and second, in the *new* model obtained by eliminating all possibilities in which φ was false, ψ is true at w. In §1.2 and §3, we will apply this truth clause in several concrete examples.

1.2. Examples

In this section, we will show that the following principles, which are *valid* for eternal sentences p, are not *schematically* valid:

1. $p \to \langle p\rangle p$

   Translation: if p is true, then after it is truly announced, it remains true.

2. $p \to \langle p\rangle K_j p$

   Translation: if p is true, then after it is truly announced, it becomes known.

3. $p \to \langle p\rangle(p \to K_j p)$

   Translation: if p is true, then after it is truly announced, it becomes known if it remains true.

4. $(p \wedge \neg K_j p) \to \langle p \wedge \neg K_j p\rangle\neg(p \wedge \neg K_j p)$

   Translation: if p is an unknown truth, then after this is truly announced, p is no longer an unknown truth.

In §3, we will show the same for the following principles:[4]

[4] Principles 5 and 6 are schematically valid for the *single-agent* language interpreted over *transitive* structures. However, they are not schematically valid for the single-agent language interpreted over arbitrary (reflexive, symmetric, Euclidean) structures or for the multi-agent language interpreted over partition models.

5. $K_j p \to \langle p \rangle K_j p$

   Translation: if p is known, then after p is truly announced, it remains known.

6. $K_j p \to \langle p \rangle (p \to K_j p)$

   Translation: if p is known, then after p is truly announced, it remains known if it remains true.

We leave it to the reader to verify that each of these principles is valid. What we will show is that for each of them, we can substitute a PKEA sentence φ for p to obtain an invalid principle. The non-schematic validity of the first two principles is the well-known problem of "unsuccessful" formulas [17, 2, 30, 18], which is also at the heart of the Muddy Children puzzle [17, §4]. Example 1 illustrates unsuccessfulness with a much-discussed style of example due to Moore [34] and Fitch [21].

Example 1 (Moorean Announcement) Suppose that agent i truly announces in the presence of agent j:

(L) "Agent j doesn't know it, but Ljubljana became the capital of an independent Slovenia in 1991."

Further suppose that j knows i to be an infallible source of information on such matters, so j accepts L. Consider the question:

Q0 After i truly announces L, is L true or false?

We can easily answer this question without using our formalism (in contrast to Examples 2 and 4 below), but let us use it as a warm up.

Let c stand for 'Ljubljana became the capital of an independent Slovenia in 1991'. Before i's announcement, j does not know whether c is true, reflected by the two uneliminated possibilities in the model $\mathcal{M}$ in Fig. 1: w_2, where c is false, and w_1 (the actual world), where c is true. As indicated by the arrows in the diagram representing agent j's epistemic relation R_j, j does not know which possibility is actual.

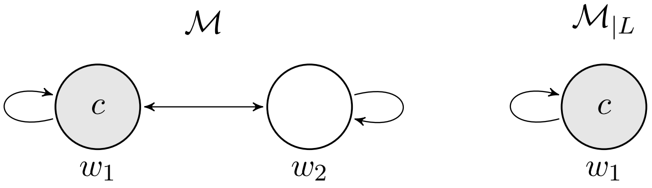

Figure 1. models for Example 1

We can translate L into our language as

$(L)\quad c \wedge \neg K_j c.$

Since L is true only at w_1 in $\mathcal{M}$,

$$[\![L]\!]^{\mathcal{M}} = \{w_1\},$$

it follows that after i's announcement, j can eliminate possibility w_2, reducing j's uncertainty to that represented by model $\mathcal{M}_{|L}$ in Fig. 1. Now j *does* know c, so

$$[\![L]\!]^{\mathcal{M}} = \emptyset,$$

which means that the answer to Q0 is

A0 After i truly announces L, L is false.

$$\mathcal{M}_{|L}, w_1 \nvDash L, \text{ so } \mathcal{M}, w_1 \nvDash \langle L \rangle L.$$

Hence the valid principle $p \to \langle p \rangle p$ is not *schematically* valid (and likewise for $p \to \langle p \rangle K_j p$). In other words, **the true announcement of a** (PKEA) **sentence may result in the sentence becoming false**.[5]

Not only is the substitution instance

$$(p \wedge \neg K_j p) \to \langle p \wedge \neg K_j p \rangle (p \wedge \neg K_j p)$$

of $p \to \langle p \rangle p$ invalid, but also

$$(p \wedge \neg K_j p) \to \langle p \wedge \neg K_j p \rangle \neg (p \wedge \neg K_j p)$$

is *valid.* As it is often put, the true announcement of a Moore sentence is "self-refuting" [5, 30]. But is this valid principle *schematically* valid? In other words, is there a φ such that if you receive the true information (from a source you know to be infallible) that "you don't know it, but φ," it can *remain true* afterward that you don't know it, but φ? Hintikka [24] remarks about sentences in the Moore schema:

> If you know that I am well informed and if I address the words ... to you, these words have a curious effect which may perhaps be called anti-performatory. You may come to know that what I say *was* true, but saying it in so many words has the effect of making what is being said false. (68-69)

[5] Of course, the true announcement of a sentence such as 'no one has ever made an announcement in this room' may result in the sentence becoming false, but it is not a PKEA sentence.

As for Hintikka's first point, that you may come to know that what he says *was* true, this can be formalized in an extension of PAL with a past operator P_φ ("before the announcement of $\varphi\ldots$") by the *schematically* valid principle $\varphi \to \langle\varphi\rangle K_i P_\varphi \varphi$ [31]. However, as for Hintikka's second point, that saying "you don't know it, but φ" has the anti-performatory effect of making what is being said false, surprisingly this is not always the case. The following puzzle provides a counterexample.

Example 2 (Puzzle of the Gifts [25]) Holding her hands behind her back, agent i walks into a room where agent j is sitting. Agent j did not see what if anything i put in her hands, and i knows this. In fact, i has gifts for j in both hands. Instead of the usual game of asking j to "pick a hand, any hand," i (deviously but) truthfully announces:

(G) Either I have a gift in my *right* hand and you don't know that, or I have gifts in *both* hands and you don't know I have a gift in my *left* hand.[6]

Let us suppose that j knows i to be an infallible source of information on such matters, so j accepts G. Consider the following questions:

Q1 After i truly announces G, does j know whether i has a gift in her left/right/both hand(s)?

Q2 After i truly announces G, is G *true*?

Q3 After i truly announces G, does j *know* G?

Q4 If 'yes' to Q2, what happens if i announces G again?

Let l stand for 'a gift is in i's left hand' and r stand for 'a gift is in i's right hand'. Before i's announcement, j has not eliminated any of the four possibilities represented by the model $\mathcal{M}$ in Fig. 2.

We can translate G into our language as

(G) $(r \wedge \neg K_j r) \vee (l \wedge r \wedge \neg K_j l)$.

First observe that

$$[\![G]\!]^{\mathcal{M}} = \{w_1, w_2\}. \tag{1}$$

Hence after i's announcement of G, j can eliminate possibilities w_3 and w_4, reducing j's uncertainty to that represented by the model $\mathcal{M}_{|G}$ in Fig. 2. Inspection of $\mathcal{M}_{|G}$ shows that the answer to Q1 is

[6] Although 'I have a gift...' is not eternal in the sense of §1.1, it could easily be eternalized.

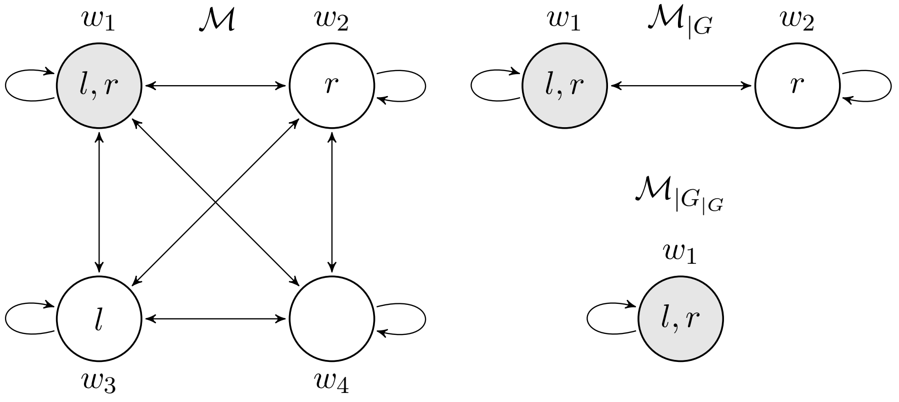

Figure 2. models for Example 2

A1 After i truly announces G, j knows that i has a gift in her right hand, but not whether i has a gift in her left hand:

$$\mathcal{M}_{|G}, w_1 \vDash K_j r \wedge \neg(K_j l \vee K_j \neg l).$$

Next observe that

$$[\![G]\!]^{\mathcal{M}_{|G}} = \{w_1\}, \tag{2}$$

which means that the answers to Q2 and Q3 are

A2 Yes—after i's announcement of G, G is true:

$$\mathcal{M}_{|G}, w_1 \vDash G, \text{ so } \mathcal{M}, w_1 \vDash \langle G \rangle G.$$

A3 No—after i's announcement of G, j does *not* know G:

$$\mathcal{M}_{|G}, w_1 \nvDash K_j G, \text{ so } \mathcal{M}, w_1 \nvDash \langle G \rangle K_j G.$$

It follows from A2 and A3 that the valid principle

$$\langle p \rangle p \to \langle p \rangle K_j p, \text{ or equivalently, } p \to \langle p \rangle (p \to K_j p),$$

is *not schematically valid*. In other words, **after the announcement of a true PKEA sentence φ by a source known to be infallible, even if φ remains true, the agent might not know it.**

We leave the answer to Q4 to the reader (see $\mathcal{M}_{|G_{|G}}$ in Fig. 2).

Suppose that instead of initially announcing G, i announces

(F) "The following is true but you don't know it: either I have a gift in my *right* hand and you don't know that, or I have gifts in *both* hands and you don't know I have a gift in my *left* hand."

We can translate F into our language as

$(F)\ \ G \wedge \neg K_j G.$

Consider the question:

Q5 After i truly announces F, is F *true*?

Given (1) above, we have

$$[\![F]\!]^{\mathcal{M}} = \{w_1, w_2\}.$$

It follows that $\mathcal{M}_{|F} = \mathcal{M}_{|G}$, and given (2) above,

$$[\![F]\!]^{\mathcal{M}_{|F}} = \{w_1\},$$

so the answer to Q5 is

A5 Yes—after i's announcement of F, F is true:

$$\mathcal{M}_{|F}, w_1 \vDash F, \text{ so } \mathcal{M}, w_1 \vDash \langle F \rangle F.$$

Given the form of F, it follows that the valid principle

$$(p \wedge \neg K_j p) \to \langle p \wedge \neg K_j p \rangle \neg (p \wedge \neg K_j p)$$

is *not schematically valid.* In other words, **true announcements of sentences in the Moore schema are not always "self-refuting."**

We have now shown the non-schematic validity of principles 1 - 4. In §3, we will show the non-schematic validity of principles 5 and 6 as well. In addition to these, there are many other examples of valid but not schematically valid principles. Noteworthy instances include

$K_i(p \to q) \to (\langle q \rangle K_i r \to \langle p \rangle K_i r)$ and

$(\langle p \rangle K_i r \wedge \langle q \rangle K_i r) \to \langle p \vee q \rangle K_i r.$

Example 2 shows that discovering there is an invalid substitution instance of a valid dynamic epistemic principle can be a non-trivial task. A natural question is whether we can give an effective procedure to make such discoveries. In §2, we will answer this question affirmatively.

1.3. Common Knowledge

Examples 1 and 2 are single-agent examples. The models are for agent j alone, with i only playing a role in our telling of the stories. However, many other examples of interesting effects in information dynamics involve multiple agents, as in the much-discussed Muddy Children puzzle [17, §4] in which repeated announcement of collective ignorance leads to knowledge. In the multi-agent setting, the elimination from a model of all possibilities in which φ was false (Def. 1) corresponds to all agents *publicly* receiving the same information. If φ is an eternal sentence p, the result is that φ becomes *commonly known* among all agents.

As standardly defined in epistemic logic, it is common knowledge that φ if and only if every agent knows φ, and every agent knows that every agent knows φ, and so on. To state when this holds in a model $\mathcal{M} = \langle W, \{R_i \mid i \in \mathsf{Agt}\}, V\rangle$, we first define the following:

- R_{Agt} is the union of the R_i relations for each $i \in \mathsf{Agt}$;
- for any relation R, R^+ is the *transitive closure* of R (the smallest transitive relation such that $R \subseteq R^+$).

Hence $R_{\mathsf{Agt}}(w) = \{v \in W \mid wR_{\mathsf{Agt}}v\}$ is the set of possibilities v such that v is consistent with *some* agent's information in w, and *everybody knows* φ at w iff φ is true throughout this set. By contrast, it is common knowledge that φ at w iff φ is true throughout the set $R^+_{\mathsf{Agt}}(w)$:

$$\mathcal{M}, w \vDash C\varphi \quad \text{iff} \quad \forall v \in R^+_{\mathsf{Agt}}(w) \colon \mathcal{M}, v \vDash \varphi.$$

In other words, $C\varphi$ is true at w iff every path from w following agents' accessibility relations ends in a possibility where φ is true. The observation above that the public announcement of p leads to common knowledge of p among all agents can be expressed by the valid principle $p \to \langle p\rangle Cp$. However, for the same reasons that principle 2 of §1.2 is not schematically valid, $p \to \langle p\rangle Cp$ is not schematically valid either.

It is a basic result of modal logic that for the purposes of evaluating what agents know at w, we can assume without loss of generality that $W = R^+_{\mathsf{Agt}}(w)$ [10, Prop. 2.6]. In other words, we may assume that our space of possibilities W includes only those possibilities that are directly or indirectly related to w. If a model satisfies this condition, then we say that w is a *root* of the model. If a model has a root, we call the model *rooted.* In §2 we will use the fact that in a rooted *partition* model (recall §1.1), R^+_{Agt} coincides with the *universal* relation on W: for all $w, v \in W$, $wR^+_{\mathsf{Agt}}v$. As a result, the common knowledge modality functions as the universal modality: $\mathcal{M}, v \vDash C\varphi$ iff $\forall u \in W \colon \mathcal{M}, u \vDash \varphi$. In a rooted *quasi-partition* model, the same is true for the defined modality $\mathbf{C}\varphi := C\varphi \wedge \varphi$. We have $\mathcal{M}, v \vDash \mathbf{C}\varphi$ iff $\forall u \in W \colon \mathcal{M}, u \vDash \varphi$.

For technical reasons, instead of adding the plain common knowledge operator C to PAL, we will add a *relativized* common knowledge operators $C^{\varphi}\psi$. The language of $\mathsf{PAL\text{-}RC}$ [9], $\mathcal{L}_{\mathsf{PAL\text{-}RC}}$, is given by

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid K_i\varphi \mid C^{\varphi}\varphi \mid \langle\varphi\rangle\varphi,$$

and the truth clause for relativized common knowledge is

$$\mathcal{M}, w \vDash C^{\varphi}\psi \text{ iff } \langle w, v\rangle \in (R_{\mathsf{Agt}} \cap (W \times [\![\varphi]\!]^{\mathcal{M}}))^{+} \text{ implies } \mathcal{M}, v \vDash \psi.$$

In other words, every path from w following agents' accessibility relations through possibilities where φ is true ends in one where ψ is true. Plain common knowledge can be defined in $\mathcal{L}_{\mathsf{PAL\text{-}RC}}$ by $C\psi := C^{\top}\psi$.

2. Decidability

We now turn to the technical part of the paper, returning to the question at the end of §1.2. For a language $\mathcal{L}$ whose set of atomic sentences is At, a substitution is any function $\sigma\colon \mathsf{At} \to \mathcal{L}$, and $\hat{\sigma}\colon \mathcal{L} \to \mathcal{L}$ is the extension such that $\hat{\sigma}(\varphi)$ is obtained from φ by replacing each $p \in \mathsf{At}(\varphi)$ by $\sigma(p)$. Abusing notation, we write $\sigma(\varphi)$ for $\hat{\sigma}(\varphi)$. A formula φ is *schematically valid* iff for all substitutions σ, the substitution instance $\sigma(\varphi)$ is valid. Finally, let the *substitution core* of PAL be the set

$$\{\varphi \in \mathcal{L}_{\mathsf{PAL}} \mid \varphi \text{ is schematically valid}\},$$

and similarly for $\mathsf{PAL\text{-}RC}$.

Question 1 (van Benthem [7]) Is the substitution core of $\mathsf{PAL\text{-}RC}$ (or PAL) *decidable*?

We will answer this question affirmatively for both $\mathsf{PAL\text{-}RC}$ and PAL over quasi-partition models (recall §1.1) with finitely many agents, as well as arbitrary models with infinitely many agents.

The idea behind the proof of these results is to show that for any φ, we can effectively construct a finite set $\mathfrak{F}(\varphi)$ of substitution instances of φ, such that if φ is not schematically valid, then there is a falsifiable substitution instance in $\mathfrak{F}(\varphi)$. We will prove that whenever φ is not schematically valid, so there is some substitution σ and model $\mathcal{M}$ with $\mathcal{M}, w \nvDash \sigma(\varphi)$, then σ can be transformed into a substitution τ such that $\tau(\varphi) \in \mathfrak{F}(\varphi)$ and $\tau(\varphi)$ is false at w in a suitable extension (on the valuation function) of $\mathcal{M}$. Therefore, to check whether φ is schematically valid, we need only check the validity of the finitely many substitution instances of φ in $\mathfrak{F}(\varphi)$, which is decidable for PAL and $\mathsf{PAL\text{-}RC}$.

2.1. Reduction

Despite the failure of uniform substitution for PAL, there is a simple axiomatization of PAL given by the axioms and rules of multi-agent epistemic logic, the rule of replacement of equivalents (from $\alpha \leftrightarrow \beta$, derive $\varphi(\alpha/p) \leftrightarrow \varphi(\beta/p)$), and the following *reduction axioms* [35]:

(i) $\langle\varphi\rangle p \leftrightarrow (\varphi \wedge p)$;

(ii) $\langle\varphi\rangle\neg\psi \leftrightarrow (\varphi \wedge \neg\langle\varphi\rangle\psi)$;

(iii) $\langle\varphi\rangle(\psi \wedge \chi) \leftrightarrow (\langle\varphi\rangle\psi \wedge \langle\varphi\rangle\chi)$;

(iv) $\langle\varphi\rangle K_i\psi \leftrightarrow (\varphi \wedge K_i(\varphi \rightarrow \langle\varphi\rangle\psi))$.

Using (i) - (iv) and replacement, any $\mathcal{L}_{\mathsf{PAL}}$ formula can be reduced to an equivalent formula in the fragment of the language without $\langle\varphi\rangle$ operators. Completeness and decidability for PAL are therefore corollaries of completeness and decidability for multi-agent epistemic logic.[7]

Similarly for $\mathsf{PAL\text{-}RC}$, using the additional reduction axiom

(v) $\langle\varphi\rangle C^{\psi}\chi \leftrightarrow (\varphi \wedge C^{\langle\varphi\rangle\psi}\langle\varphi\rangle\chi)$,

any $\mathcal{L}_{\mathsf{PAL\text{-}RC}}$ formula can be reduced to an equivalent formula without dynamic operators. Hence an axiomatization for $\mathsf{PAL\text{-}RC}$ may be obtained from (i) - (v), the rule of replacement, and the axioms and rules for multi-agent epistemic logic with relativized common knowledge [9]. Since the latter system is decidable, so is $\mathsf{PAL\text{-}RC}$ by the reduction.

Although by using (i) - (v) we can reduce any φ to an equivalent φ' containing no dynamic operators, there is no guarantee that φ and φ' will be *schematically* equivalent (i.e., $\varphi \leftrightarrow \varphi'$ schematically valid), since (i) itself is not schematically valid. However, we can at least reduce any φ to a schematically equivalent φ' of a certain simple form.

Definition 2 (Simple Formulas) The set of *simple* formulas is generated by the grammar

$$\varphi ::= p \mid \neg\varphi \mid (\varphi \wedge \varphi) \mid K_i\varphi \mid C^{\varphi}\varphi \mid \langle\varphi\rangle p,$$

where $p \in \mathsf{At}$ and $i \in \mathsf{Agt}$.

Since only atomic sentences may occur after dynamic operators in simple formulas, there can be no consecutive occurrences of dynamic operators, though nesting is fine. Dealing with only simple formulas will be convenient (though not essential) in our proof below. That

[7] For alternative axiomatizations of PAL, see the recent study by Wang [40].

we only need to deal with such formulas follows from an interesting schematically valid principle relating consecutive occurrences of dynamic operators to nested occurrences: $\langle p\rangle\langle q\rangle r \leftrightarrow \langle\langle p\rangle q\rangle r$ [16, 40].

Proposition 1 (Simple Reduction) For every formula $\varphi \in \mathcal{L}_{\mathsf{PAL\text{-}RC}}$, there is a schematically equivalent simple formula φ'.

Proof By induction on φ, using the schematically valid reduction axioms (ii) - (v) and the schematic validity $\langle p\rangle\langle q\rangle r \leftrightarrow \langle\langle p\rangle q\rangle r$. □

2.2. Transforming Substitutions

Fix a formula φ in $\mathcal{L}_{\mathsf{PAL}}$ or $\mathcal{L}_{\mathsf{PAL\text{-}RC}}$. By Proposition 1, we may assume that φ is simple. Suppose that for some substitution σ and quasi-partition $\mathcal{M} = \langle W, \{R_i \mid i \in \mathsf{Agt}\}, V\rangle$, $\mathcal{M}, w \nvDash \sigma(\varphi)$.[8] We will now show how to construct a special substitution τ from σ and a model $\mathcal{N}$ from $\mathcal{M}$ such that $\mathcal{N}, w \nvDash \tau(\varphi)$, as discussed before §2.1. Whether φ is in $\mathcal{L}_{\mathsf{PAL}}$ or $\mathcal{L}_{\mathsf{PAL\text{-}RC}}$, the resulting formula $\tau(\varphi)$ will be in $\mathcal{L}_{\mathsf{PAL\text{-}RC}}$. However, in §2.3 we will obtain substitution instances in $\mathcal{L}_{\mathsf{PAL}}$.

To construct $\tau(p)$ for a given $p \in \mathsf{At}(\varphi)$, let $D_1, \dots, D_m$ be the sequence of all formulas D_i such that $\langle D_i\rangle p$ occurs in φ, and let $D_0 := \top$. For $0 \leq i, j \leq m$, if $[\![\sigma(D_i)]\!]^{\mathcal{M}} = [\![\sigma(D_j)]\!]^{\mathcal{M}}$, then delete one of D_i or D_j from the list (but never D_0), until there is no such pair. Call the resulting sequence $A_0, \dots, A_n$, and define

$$s(i) = \{j \mid 0 \leq j \leq n \text{ and } [\![\sigma(A_j)]\!]^{\mathcal{M}} \subsetneq [\![\sigma(A_i)]\!]^{\mathcal{M}}\}.$$

Extend the language with new atoms $p_0, \dots, p_n$ and $a_0, \dots, a_n$, and define $\tau(p) = \kappa_0 \wedge \ldots \wedge \kappa_n$ such that

$$\kappa_i := p_i \vee \bigvee_{0\leq j\leq n,\, j\neq i} \Big(\mathbf{C}a_j \wedge \bigwedge_{0\leq k\leq n,\, k\in s(j)} \neg\mathbf{C}a_k\Big),$$

where $\mathbf{C}\varphi := C\varphi \wedge \varphi$. As noted in §1.3, we may assume without loss of generality that $\mathcal{M}$ is rooted at w, so that the $\mathbf{C}$ modality functions as the *universal* modality in $\mathcal{M}$, given that $\mathcal{M}$ is a quasi-partition. This is an important point, discussed further in §3.

For other $q \in \mathsf{At}(\varphi)$, extend the language and construct $\tau(q)$ analogously. Having thereby extended the language for each $p \in \mathsf{At}(\varphi)$, extend the valuation V to V' such that for each $p \in \mathsf{At}(\varphi)$, $V'(p) = V(p)$, and for the new atoms:

(a) $V'(p_i) = [\![\sigma(p)]\!]^{\mathcal{M}|\sigma(A_i)}$;

[8] Later we will lift the assumption that $\mathcal{M}$ is a quasi-partition, when considering a language with infinitely many agents.

(b) $V'(a_i) = [\![\sigma(A_i)]\!]^{\mathcal{M}}$.

Let $\mathcal{N} = \langle W, \{R_i \mid i \in \mathsf{Agt}\}, V' \rangle$ be the extension of $\mathcal{M}$ with V', so

($\mathbf{a}$) $[\![p_i]\!]^{\mathcal{N}} = [\![\sigma(p)]\!]^{\mathcal{M}_{|\sigma(A_i)}}$;

($\mathbf{b}$) $[\![a_i]\!]^{\mathcal{N}} = [\![\sigma(A_i)]\!]^{\mathcal{M}}$.

Note that it follows from ($\mathbf{a}$) and Definition 1 that

($\mathbf{c}$) $[\![p_i]\!]^{\mathcal{N}} = [\![\sigma(\langle A_i \rangle p)]\!]^{\mathcal{M}}$.

Using these facts, we will show that in $\mathcal{N}$, $\tau(p)$ has the same extension as $\sigma(p)$ after update with any $\sigma(A_i)$, which has the same extension as $\tau(A_i)$. It will follow that $\mathcal{N}, w \nvDash \tau(\varphi)$ given $\mathcal{M}, w \nvDash \sigma(\varphi)$.

Lemma 1 For all $0 \leq i \leq n$,

$$[\![\tau(p)]\!]^{\mathcal{N}_{|a_i}} = [\![p_i]\!]^{\mathcal{N}}.$$

Proof We first show that for $0 \leq i, j \leq n$, $i \neq j$:

1. $[\![\kappa_i]\!]^{\mathcal{N}_{|a_i}} = [\![p_i]\!]^{\mathcal{N}_{|a_i}}$;
2. $[\![\kappa_j]\!]^{\mathcal{N}_{|a_i}} = [\![a_i]\!]^{\mathcal{N}_{|a_i}}$ $(= W_{|a_i})$.

For 1, we claim that given $i \neq j$,

$$[\![\mathbf{C}a_j \wedge \bigwedge_{0 \leq k \leq n,\ k \in s(j)} \neg \mathbf{C}a_k]\!]^{\mathcal{N}_{|a_i}} = \emptyset.$$

By construction of the sequence $A_0, \dots, A_n$ for p and ($\mathbf{b}$), $[\![a_j]\!]^{\mathcal{N}} \neq [\![a_i]\!]^{\mathcal{N}}$. We consider two cases. First, if $[\![a_i]\!]^{\mathcal{N}} \not\subsetneq [\![a_j]\!]^{\mathcal{N}}$, then $[\![\mathbf{C}a_j]\!]^{\mathcal{N}_{|a_i}} = \emptyset$. Second, if $[\![a_i]\!]^{\mathcal{N}} \subsetneq [\![a_j]\!]^{\mathcal{N}}$, then by ($\mathbf{b}$) and the definition of s, $i \in s(j)$. Then since a_i is propositional, $[\![\neg \mathbf{C}a_i]\!]^{\mathcal{N}_{|a_i}} = \emptyset$. In either case the claim holds, so $[\![\kappa_i]\!]^{\mathcal{N}_{|a_i}} = [\![p_i]\!]^{\mathcal{N}_{|a_i}}$ given the structure of κ_i.

For 2, κ_j contains as a disjunct

$$\mathbf{C}a_i \wedge \bigwedge_{0 \leq k \leq n,\ k \in s(i)} \neg \mathbf{C}a_k.$$

Since a_i is propositional, $[\![\mathbf{C}a_i]\!]^{\mathcal{N}_{|a_i}} = W_{|a_i}$. By definition of s and ($\mathbf{b}$), for all $k \in s(i)$, $[\![a_k]\!]^{\mathcal{N}} \subsetneq [\![a_i]\!]^{\mathcal{N}}$, which gives $[\![\neg \mathbf{C}a_k]\!]^{\mathcal{N}_{|a_i}} = W_{|a_i}$. Hence $[\![\kappa_j]\!]^{\mathcal{N}_{|a_i}} = W_{|a_i}$.

Given the construction of τ, 1 and 2 imply:

$$[\![\tau(p)]\!]^{\mathcal{N}_{|a_i}} = [\![\kappa_i]\!]^{\mathcal{N}_{|a_i}} \cap \bigcap_{j \neq i} [\![\kappa_j]\!]^{\mathcal{N}_{|a_i}} = [\![p_i]\!]^{\mathcal{N}_{|a_i}} \cap [\![a_i]\!]^{\mathcal{N}_{|a_i}} = [\![p_i]\!]^{\mathcal{N}},$$

where the last equality holds because $[\![p_i]\!]^{\mathcal{N}} \subseteq [\![a_i]\!]^{\mathcal{N}}$, which follows from ($\mathbf{a}$) and ($\mathbf{b}$). □

Lemma 2 For all simple subformulas χ of φ,

$$\llbracket \tau(\chi) \rrbracket^{\mathcal{N}} = \llbracket \sigma(\chi) \rrbracket^{\mathcal{M}}.$$

Proof By induction on χ. For the base case, we must show $\llbracket \tau(p) \rrbracket^{\mathcal{N}} = \llbracket \sigma(p) \rrbracket^{\mathcal{M}}$ for $p \in \mathsf{At}(\varphi)$. By construction of the sequence $A_0, \dots, A_n$ for p, $A_0 = \top$, so $\llbracket \sigma(A_0) \rrbracket^{\mathcal{M}} = W$. Then by (**b**), $\llbracket a_0 \rrbracket^{\mathcal{N}} = W$, and hence

$$\begin{aligned} \llbracket \tau(p) \rrbracket^{\mathcal{N}} &= \llbracket \tau(p) \rrbracket^{\mathcal{N}|a_0} \\ &= \llbracket p_0 \rrbracket^{\mathcal{N}} && \text{by Lemma 1} \\ &= \llbracket \sigma(p) \rrbracket^{\mathcal{M}|\sigma(A_0)} && \text{by } (\mathbf{a}) \\ &= \llbracket \sigma(p) \rrbracket^{\mathcal{M}}. \end{aligned}$$

The boolean cases are straightforward. Next, we must show $\llbracket \tau(K_k\varphi) \rrbracket^{\mathcal{N}}$ $\llbracket \sigma(K_k\varphi) \rrbracket^{\mathcal{M}}$. For the inductive hypothesis, we have $\llbracket \tau(\varphi) \rrbracket^{\mathcal{N}} = \llbracket \sigma(\varphi) \rrbracket^{\mathcal{M}}$, so

$$\begin{aligned} \llbracket \tau(K_k\varphi) \rrbracket^{\mathcal{N}} &= \llbracket K_k\tau(\varphi) \rrbracket^{\mathcal{N}} \\ &= \{w \in W \mid R_k(w) \subseteq \llbracket \tau(\varphi) \rrbracket^{\mathcal{N}}\} \\ &= \{w \in W \mid R_k(w) \subseteq \llbracket \sigma(\varphi) \rrbracket^{\mathcal{M}}\} \\ &= \llbracket K_k\sigma(\varphi) \rrbracket^{\mathcal{M}} \\ &= \llbracket \sigma(K_k\varphi) \rrbracket^{\mathcal{M}}. \end{aligned}$$

Similar reasoning applies in the case of $C^{\varphi}\psi$.

Finally, we must show $\llbracket \tau(\langle D_i \rangle p) \rrbracket^{\mathcal{N}} = \llbracket \sigma(\langle D_i \rangle p) \rrbracket^{\mathcal{M}}$. For the inductive hypothesis, $\llbracket \tau(D_i) \rrbracket^{\mathcal{N}} = \llbracket \sigma(D_i) \rrbracket^{\mathcal{M}}$. By construction of the sequence $A_0, \dots, A_n$ for $p \in \mathsf{At}(\varphi)$, there is some A_j such that

$$(\star) \quad \llbracket \sigma(D_i) \rrbracket^{\mathcal{M}} = \llbracket \sigma(A_j) \rrbracket^{\mathcal{M}}.$$

Therefore,

$$\begin{aligned} \llbracket \tau(D_i) \rrbracket^{\mathcal{N}} &= \llbracket \sigma(A_j) \rrbracket^{\mathcal{M}} \\ &= \llbracket a_j \rrbracket^{\mathcal{N}} && \text{by } (\mathbf{b}), \end{aligned}$$

and hence

$$\begin{aligned} \llbracket \tau(\langle D_i \rangle p) \rrbracket^{\mathcal{N}} &= \llbracket \langle \tau(D_i) \rangle \tau(p) \rrbracket^{\mathcal{N}} \\ &= \llbracket \langle a_j \rangle \tau(p) \rrbracket^{\mathcal{N}} \\ &= \llbracket \tau(p) \rrbracket^{\mathcal{N}|a_j} \\ &= \llbracket p_j \rrbracket^{\mathcal{N}} && \text{by Lemma 1} \\ &= \llbracket \sigma(\langle A_j \rangle p) \rrbracket^{\mathcal{M}} && \text{by } (\mathbf{c}) \\ &= \llbracket \sigma(\langle D_i \rangle p) \rrbracket^{\mathcal{M}} && \text{by } (\star). \end{aligned}$$

The proof by induction is complete. □

Corollary 1 If $\mathcal{M}, w \nvDash \sigma(\varphi)$, then $\mathcal{N}, w \nvDash \tau(\varphi)$.

Proof Immediate from Lemma 2. □

Before using the results of this section to obtain our main theorems in §2.3, let us work out an example for the sake of concreteness.

Example 3 Let $\varphi := p \to \langle p \rangle p$, $\sigma(p) = p \wedge \neg K_i p$, and $\mathcal{M}$ be the model in Fig. 3. As noted in Example 1, $\mathcal{M}, w_1 \vDash \varphi$ but $\mathcal{M}, w_1 \nvDash \sigma(\varphi)$. Now let us obtain τ and $\mathcal{N}$ as in the proof above. First, we have

$A_0 := \top$, so $\sigma(A_0) = \top$;

$A_1 := p$, so $\sigma(A_1) = p \wedge \neg K_i p$.

Given our definition of the function s by

$$s(i) = \{j \mid 0 \leq j \leq n \text{ and } [\![\sigma(A_j)]\!]^{\mathcal{M}} \subsetneq [\![\sigma(A_i)]\!]^{\mathcal{M}}\},$$

we have $s(0) = \{1\}$ and $s(1) = \emptyset$.

Next we introduce new atoms a_0, a_1, p_0, and p_1, such that

$[\![a_0]\!]^{\mathcal{N}} = [\![\sigma(A_0)]\!]^{\mathcal{M}} = \{w_1, w_2\}$;

$[\![a_1]\!]^{\mathcal{N}} = [\![\sigma(A_1)]\!]^{\mathcal{M}} = \{w_1\}$;

$[\![p_0]\!]^{\mathcal{N}} = [\![\sigma(p)]\!]^{\mathcal{M}|\sigma(A_0)} = \{w_1\}$;

$[\![p_1]\!]^{\mathcal{N}} = [\![\sigma(p)]\!]^{\mathcal{M}|\sigma(A_1)} = \emptyset$.

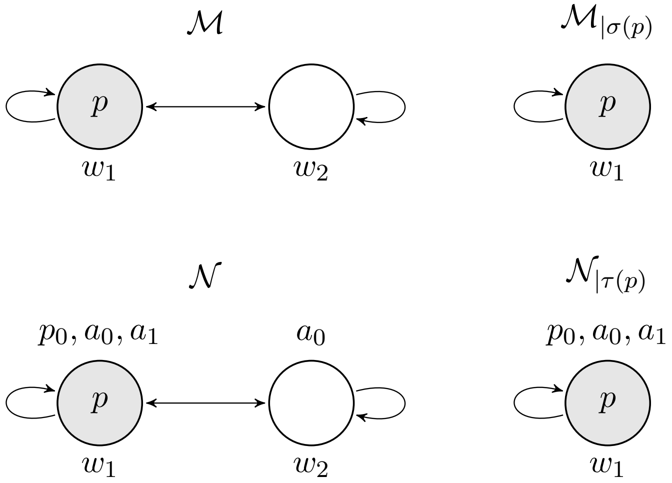

Figure 3. models for Example 3

Recall that we define $\tau(p) = \kappa_0 \wedge \ldots \wedge \kappa_n$ such that

$$\kappa_i := p_i \vee \bigvee_{0 \leq j \leq n,\, j \neq i} \Big(\mathbf{C}a_j \wedge \bigwedge_{0 \leq k \leq n,\, k \in s(j)} \neg \mathbf{C}a_k\Big),$$

which in this case gives

$$\tau(p) := (p_0 \vee \mathbf{C}a_1) \wedge (p_1 \vee (\mathbf{C}a_0 \wedge \neg \mathbf{C}a_1)).$$

which is equivalent to

$$(p_0 \vee K_i a_1) \wedge (p_1 \vee (K_i a_0 \wedge \neg K_i a_1))$$

given that R_i is the universal relation in this case (see the end of §2.3).

Finally, observe that $[\![\tau(p)]\!]^{\mathcal{N}} = \{w_1\}$ and $[\![\tau(p)]\!]^{\mathcal{N}|_{\tau(p)}} = \emptyset$, so we have $\mathcal{N}, w_1 \nvDash \tau(\varphi)$, as desired.

2.3. Proof of Decidability

We can now state the first of our three main results in Theorem 1.

As usual, a *frame* is a pair $\mathcal{F} = \langle W, \{R_i \mid i \in \mathsf{Agt}\}\rangle$; a *quasi-partition* (resp. *partition*) frame is a frame in which each R_i is transitive and Euclidean (resp. an equivalence relation); a formula φ is *valid on* $\mathcal{F} = \langle W, \{R_i \mid i \in \mathsf{Agt}\}\rangle$ iff for all models $\mathcal{M} = \langle W, \{R_i \mid i \in \mathsf{Agt}\}, V\rangle$ and $w \in W$, $\mathcal{M}, w \vDash \varphi$; for language $\mathcal{L}$, the *$\mathcal{L}$-theory* of a class $\mathbb{F}$ of frames is the set of all formulas in $\mathcal{L}$ that are valid on all frames in $\mathbb{F}$; and finally, the *substitution core* of a theory is the set of formulas in the theory all of whose substitution instances are also in the theory.

Theorem 1 (Decidability for $\mathcal{L}_{\mathsf{PAL\text{-}RC}}$ over Quasi-Partitions) If the $\mathcal{L}_{\mathsf{PAL\text{-}RC}}$-theory of a class of quasi-partition frames is decidable, then the substitution core of the theory is decidable.

Proof Suppose we are given a formula $\varphi \in \mathcal{L}_{\mathsf{PAL\text{-}RC}}$ with atomic sentences $p^1, \dots, p^t$ and m occurrences of dynamic operators. If for each $l \leq t$, we choose a number $n_l \leq m$, introduce the new atoms $p^l_0, \dots, p^l_{n_l}$ and $a^l_0, \dots, a^l_{n_l}$, and choose a function $s_l : \{0, \dots, n_l\} \to \wp(\{0, \dots, n_l\})$, then we can define a substitution $\tau_{n_1,s_1;\dots;n_t,s_t}$ such that for all $l \leq t$, $\tau_{n_1,s_1;\dots;n_t,s_t}(p^l) = \kappa^l_0 \wedge \dots \wedge \kappa^l_{n_l}$, where for all $i \leq n_l$,

$$\kappa^l_i := p^l_i \vee \bigvee_{0 \leq j \leq n_l,\, j \neq i} \Bigg(\mathbf{C}a^l_j \wedge \bigwedge_{0 \leq k \leq n_l,\; k \in s_l(j)} \neg \mathbf{C}a^l_k\Bigg),$$

and for all $q \notin \mathsf{At}(\varphi)$, $\tau_{n_1,s_1;\dots;n_t,s_t}(q) = \top$. Now consider the set of all such substitutions, varying for each $l \leq t$ the choice of n_l and s_l:

$$\mathfrak{T}(\varphi) = \{\tau_{n_1,s_1;\dots;n_t,s_t} \mid \forall l \leq t : n_l \leq m, s_l : \{0, \dots, n_l\} \to \wp(\{0, \dots, n_l\})\}.$$

Clearly $\mathfrak{T}(\varphi)$ is finite. Given a class $\mathbb{F}$ of quasi-partition frames, we claim that φ is schematically valid[9] over $\mathbb{F}$ iff for every $\tau \in \mathfrak{T}(\varphi)$,

[9] We say that φ is schematically valid *over a class* $\mathbb{F}$ of frames iff for all substitutions σ, $\sigma(\varphi)$ true at all points in all models based on frames in $\mathbb{F}$.

$\tau(\varphi)$ is valid over $\mathbb{F}$. The left-to-right direction is immediate from the definition of schematic validity. For the right-to-left direction, if φ is not schematically valid over $\mathbb{F}$, then as shown in §2.2, there exists a $\tau \in \mathfrak{T}(\varphi)$ such that $\tau(\varphi)$ is not valid over $\mathbb{F}$. (Note that the model $\mathcal{N}$ of Corollary 1 is based on the same frame as the initial model $\mathcal{M}$ in §2.2.) Finally, by the assumption that the theory of $\mathbb{F}$ is decidable, we have a decision procedure for $\mathsf{PAL\text{-}RC}$ schematic validity over $\mathbb{F}$: check the validity over $\mathbb{F}$ of $\tau(\varphi)$ for each of the finitely many $\tau \in \mathfrak{T}(\varphi)$. □

As a sample application of Theorem 1, since the $\mathcal{L}_{\mathsf{PAL\text{-}RC}}$-theory of the class of all n-agent partition frames is decidable [9], we have the following result, which answers van Benthem's question from §1.

Corollary 2 The substitution core of $\mathsf{PAL\text{-}RC}\text{-}\mathbf{S5}_n$, the $\mathcal{L}_{\mathsf{PAL\text{-}RC}}$-theory of all n-agent partition frames, is decidable.

Recall from §2.2 that we assumed $\mathsf{PAL\text{-}RC}$ over *quasi-partition* models so that the $\mathbf{C}$ modality functions as the universal modality (also see §1.3). In a *finite* quasi-partition model, we can simulate the universal modality without the $\mathbf{C}$ modality, so we have the following result.

Theorem 2 (Decidability for $\mathcal{L}_{\mathsf{PAL}}$ over Quasi-Partitions) If the $\mathcal{L}_{\mathsf{PAL\text{-}RC}}$-theory of a class $\mathbb{F}$ of quasi-partition frames has the effective finite model property,[10] then the substitution core of the $\mathcal{L}_{\mathsf{PAL}}$-theory of $\mathbb{F}$ is decidable.

Proof Given $\varphi \in \mathcal{L}_{\mathsf{PAL}}$, first decide whether there is a falsifiable substitution instance of φ in $\mathcal{L}_{\mathsf{PAL\text{-}RC}}$, using the procedure of Theorem 1. If there is none, then there is no falsifiable substitution instance of φ in $\mathcal{L}_{\mathsf{PAL}}$, so φ is in the substitution core of the $\mathcal{L}_{\mathsf{PAL}}$-theory of $\mathbb{F}$. If there such a substitution instance $\tau(\varphi)$, then by the effective finite model property, we can effectively find a finite model $\mathcal{M}$ for which $\mathcal{M}, w \nvDash \tau(\varphi)$. Since the C operator occurs in $\tau(p)$, we have $\tau(\varphi) \in \mathcal{L}_{\mathsf{PAL\text{-}RC}}$. But since $\varphi \in \mathcal{L}_{\mathsf{PAL}}$, we may now obtain a substitution τ' with $\tau'(\varphi) \in \mathcal{L}_{\mathsf{PAL}}$ such that $\mathcal{M}, w \nvDash \tau'(\varphi)$. We use the fact that since $\mathcal{M}$ is finite, we can define the formula $C\alpha$ in $\mathcal{M}$ by $E^{|\mathcal{M}|}\alpha$, where

$$E^1\alpha := \alpha \wedge \bigwedge_{i \in \mathsf{Agt}} K_i\alpha \text{ and } E^{n+1}\alpha := \alpha \wedge EE^n\alpha.$$

Modify τ to τ' by replacing all occurrences of $C\alpha$ in $\tau(\varphi)$ by $E^{|\mathcal{M}|}\alpha$. It is straightforward to verify that $\mathcal{M}, w \nvDash \tau'(\varphi)$ given $\mathcal{M}, w \nvDash \tau(\varphi)$. □

[10] We say that the theory of a class $\mathbb{F}$ of frames has the effective finite model property iff there is an effective procedure such that given a formula φ that is not in the theory, the procedure outputs a finite model in which φ is false.

As a sample application of Theorem 2, since the $\mathcal{L}_{\mathsf{PAL\text{-}RC}}$-theory of the class of all n-agent partition frames has the effective finite model property by the completeness proof of [9], we have the following result.

Corollary 3 The substitution core of PAL-$\mathbf{S5}_n$, the $\mathcal{L}_{\mathsf{PAL}}$-theory of all n-agent partition frames, is decidable.

In addition to simulating the universal modality as above, we can simply interpret a K_j modality as the universal modality in a given model, provided K_j does not occur in our formula φ. While the method of simulating the universal modality using common knowledge requires quasi-partition frames, by using the method of reinterpretation we can extend our results to *any* frame class, provided there are infinitely many agent modalities K_j in our language ($|\mathsf{Agt}| \geq \omega$). In this case, suppose we have any model $\mathcal{M}$ and substitution σ such that $\mathcal{M}, w \nvDash \sigma(\varphi)$. (As in §2.2, we can assume that w is the root of $\mathcal{M}$.) For any K_j operator not occurring in φ, let $\mathcal{M}_j$ be the extension of $\mathcal{M}$ in which R_j is the universal relation, so wR_jv for all $w, v \in W$. The proof in §2.2 now applies starting with $\mathcal{M}_j$, only we modify τ to τ' by replacing all occurrences of $\mathbf{C}$ in $\tau(\varphi)$ by K_j. Instead of Corollary 1, we have the fact that $\mathcal{N}_j, w \nvDash \tau'(\varphi)$ if $\mathcal{M}, w \nvDash \sigma(\varphi)$. Then the proof of the following result is the same as that of Theorem 1, but with K_j in place of $\mathbf{C}$.

Theorem 3 (General Decidability for $\mathcal{L}^{\omega}_{\mathsf{PAL}}$ and $\mathcal{L}^{\omega}_{\mathsf{PAL\text{-}RC}}$)
If the theory of a class of frames in the language $\mathcal{L}^{\omega}_{\mathsf{PAL}}$ (or $\mathcal{L}^{\omega}_{\mathsf{PAL\text{-}RC}}$) of PAL (or $\mathsf{PAL\text{-}RC}$) with infinitely many agents is decidable, then the substitution core of the theory is decidable.

3. Discussion

We began in §1 with the picture of a space of related possibilities representing agents' uncertainty about the world, about what information other agents possess, and even about their own information states. Using the formalization of this picture in dynamic epistemic logic, we have explored the relation between information dynamics and failures of uniform substitution, as exhibited by Examples 1 and 2.

The proof in §2.2 shows the kind of semantic and syntactic structure that is sufficient to induce failures of uniform substitution. What is remarkable is how simple this structure can be. Although the special substitution τ in §2.2 involves common knowledge, we have seen in §2.3 that what is essential to the proof is that we have a modality for the simplest non-empty relation on the space W of possibilities: the universal relation. The universal relation is the epistemic accessibility relation for an agent who is maximally ignorant about her location in the space of possibilities W. She knows that her actual world is somewhere in W, but nothing more. If we can describe the knowledge and ignorance of such an agent, at just one level of iteration, then we can always "fill in the blanks" to obtain an invalid substitution instance of a non-schematically valid dynamic epistemic principle. No complicated iterations of agents' knowledge operators are necessary.

In models for the knowledge of a fully introspective agent j (single-agent rooted partition models), j's epistemic accessibility relation R_j is already the universal relation, so there is no need to consider a new agent in order to decide the schematic validity of a principle φ about j. In models for the knowledge of many fully introspective agents (multi-agent rooted partition models), the common knowledge relation R^+_{Agt} (recall §1.3) is the universal relation, so again there is no need to bring in a new agent. However, in models for the knowledge of one or more agents who are not fully introspective, the universal relation does not necessarily coincide with any individual agent's relation or with the common knowledge relation. This is why we introduced a new agent whose accessibility relation is the universal relation at the end of §2.3.

If we put no bound on the number of agents, then we can always introduce a special new agent in this way.[11] However, it is also an interesting question whether a valid principle has an invalid substitution instance that only describes the knowledge of agents *whose knowledge is described by the original principle.* In other words, where

$$\mathsf{Agt}(\varphi) = \{i \in \mathsf{Agt} \mid K_i \text{ occurs in } \varphi\},$$

the question is whether there is an invalid substitution instance $\sigma(\varphi)$ of the valid φ such that $\mathsf{Agt}(\sigma(\varphi)) = \mathsf{Agt}(\varphi)$. In the single-agent case, the question is whether a valid single-agent principle has an invalid *single-agent* substitution instance. While we have shown how we can always answer this question for a fully introspective agent, it remains to be seen how to answer the question in general for an agent without full introspection. Example 4 below shows the interest of this question.

In Examples 1 and 2, we considered an introspective agent j whose accessibility relation R_j is an equivalence relation. In the next example, inspired by Williamson's [41] arguments against the positive introspection principle $K_j\varphi \to K_jK_j\varphi$, we consider an agent with a *non-transitive* accessibility relation. The example shows that the following principles (recall §1.2) that are schematically valid over single-agent quasi-partitions are not schematically valid over a wider class of models:

[11] We are asumming, as usual in epistemic logic, that the model class is not restricted so as to prohibit agents from having universal accessibility relations.

5. $K_j p \to \langle p \rangle K_j p$

   Translation: if p is known, then after p is truly announced, it remains known.

6. $K_j p \to \langle p \rangle (p \to K_j p)$

   Translation: if p is known, then after p is truly announced, it remains known if it remains true.

If we allow ourselves to substitute for p some description of the knowledge of a *different* agent k, then we can demonstrate the non-schematic validity of these principles in ways similar to those in §1.2: for example, for principle 5, substitute the Moore sentence $p \wedge \neg K_k p$ for p. However, if we require that $\mathsf{Agt}(\sigma(\varphi)) = \mathsf{Agt}(\varphi)$, then demonstrating the non-schematic validity of these principles requires more ingenuity.

Example 4 (The Heights [25]) Agent i asks agent j to estimate the height of an object at a distance after offering the following hint: if the object's height is 10 units or fewer, then its value is one of $1, 2, \ldots, 10$, while if the object's height is above 10 units, then its value is one of $10.5, 11, 11.5, \ldots, 20$. Suppose that given j's limited powers of discrimination, for any value n, if an object at the given distance from j is of height n, then whatever j believes about its height, j does not know that it is not of height $n - .5$ or $n + .5$; but if the height is at the midpoint of $n = 10$, she can always tell that it is not $n - 1$. Further suppose that j knows all of this about herself. Finally, suppose that the actual height of the object is 9, but j mistakenly believes it is 10.

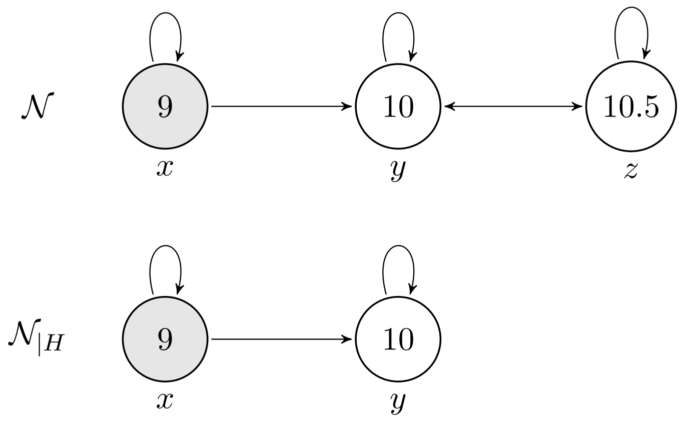

Figure 4. models for Example 4

Agent i asks agent j, "What do you think the object's height is?"

Agent j replies, "I believe it's 10, but given my limited discrimination, I must admit that I don't *know* it's 10, because even if I'm right and I know it isn't 9, I can't know it isn't 10.5 if it's as close as 10."

We can represent j's belief as follows:[12]

$$B_j(10 \wedge \neg K_j 10 \wedge K_j \neg 9 \wedge \neg K_j \neg 10.5).$$

Agent j's uncertainty is represented by the model $\mathcal{N}$ in Fig. 4. The important feature of $\mathcal{N}$ is the *non-transitivity* of the epistemic accessibility relation. The possibility y in which the answer to i's question is 10 (in which case j can tell it isn't 9) is consistent with what j knows in the actual world x, for in x she falsely believes she is in y. Moreover, if she *were* in y, then the possibility z in which the answer is 10.5 would be consistent with her knowledge, given what we have assumed to be known about her limited discrimination. However, in the actual world x in which the answer is 9, she can tell that it isn't 10.5, so z is not consistent with what she knows in x. The catch is that although in x she knows the answer isn't 10.5, she doesn't know she knows this.[13]

While agent j's belief represented above is false in the actual world x, there is something else she believes that is true in x:

(H) $(9 \vee 10) \wedge \neg K_j 10.$

Indeed, j knows H:

$$\mathcal{N}, x \vDash K_j H.$$

Now suppose that i tells j what j already knows:

(H) Agent j doesn't know the answer is 10, but it's 9 or 10.

Consider the questions

Q6 After i tells j what j already knows, H, is H still true?

Q7 After i tells j what j already knows, H, does j still know H?

First observe that

$$[\![H]\!]^{\mathcal{N}} = \{x, y\}.$$

Hence after i's announcement of H, j can eliminate possibility z, reducing j's uncertainty to that represented by the model $\mathcal{N}_{|H}$ in Fig. 4. Next observe that

$$[\![H]\!]^{\mathcal{N}} = \{x\},$$

which means that the answers to Q6 and Q7 are

[12] To represent j's beliefs in the model $\mathcal{N}$ of Fig. 4, we can simply add another relation (as if for another agent) $\{\langle x, y\rangle, \langle y, y\rangle, \langle z, y\rangle\}$, reflecting that j believes that possibility y is the actual situation.

[13] Here we assume that in x, j's belief that the answer isn't 10.5 can constitute knowledge even if she believes that she doesn't know it isn't 10.5. The example could be complicated to remove this assumption, but it would not affect our main point.

A6 Yes—after i tells j what j knows, H, H is true:

$$\mathcal{N}_{|H}, x \vDash H, \text{ so } \mathcal{N}, x \vDash K_j H \wedge \langle H \rangle H.$$

A7 No—after i tells j what j knows, H, j doesn't know H:

$$\mathcal{N}_{|H}, x \vDash \neg K_j H, \text{ so } \mathcal{N}, x \vDash K_j H \wedge \langle H \rangle \neg K_j H.$$

The reason j no longer knows H, though it is still true, is that *her knowing that the answer is* 10 (as in y in $\mathcal{N}_{|H}$) is now consistent with what she knows in x. Note that the important feature of $\mathcal{N}_{|H}$ for this result is that the epistemic accessibility relation is not *symmetric*.

It follows from the previous observations that the valid principles

$$K_j p \to \langle p \rangle K_j p \text{ and even } K_j p \to \langle p \rangle (p \to K_j p)$$

are *not schematically valid*. In other words, **as a result of being told φ by a source of information known to be infallible, an agent who initially knows φ may not know that φ is true afterward,**[14] **even though it is.**

Example 4 shows that in the case of agents without full introspection, searching for an invalid *single-agent* substitution instance of a valid single-agent principle can lead to the discovery of interesting epistemic phenomena, which we may miss if we settle for an invalid *multi-agent* substitution instance. More generally, searching for an invalid substitution instance $\sigma(\varphi)$ of φ where $\mathsf{Agt}(\sigma(\varphi)) = \mathsf{Agt}(\varphi)$ may lead to discoveries that we may miss if we allow $\mathsf{Agt}(\varphi) \subsetneq \mathsf{Agt}(\sigma(\varphi))$. Moreover, if we restrict our attention to finitely many agents, then we will be forced to find substitutions for which $\mathsf{Agt}(\sigma(\varphi)) = \mathsf{Agt}(\varphi)$, since there will be formulas φ for which $\mathsf{Agt}(\varphi) = \mathsf{Agt}$. This observation motivates the following question, which we leave as an open problem:

- Does an analogue of Theorem 3 hold for finitely many agents?

Example 4 shows not only that we may find an invalid substitution instance of a single-agent principle without introducing any new agent, but also that we may find an invalid substitution instance with a different syntactic structure than that of the substitution instances produced by our proof in §2.2. This observation motivates another question:

[14] Another way to see that $K_j p \to \langle p \rangle K_j p$ and even $K_j p \to \langle K_j p \rangle K_j p$ are not schematically valid (even over models with symmetric accessibility relations, unlike R_j in Fig. 4) is to observe that $K_j(p \wedge \neg K_j K_j p) \to \langle p \wedge \neg K_j K_j p \rangle \neg K_j (p \wedge \neg K_j K_j p)$ and $K_j(p \wedge \neg K_j K_j p) \to \langle K_j(p \wedge \neg K_j K_j p) \rangle \neg K_j(p \wedge \neg K_j K_j p)$ are *valid* and their antecedents are satisfiable without transitivity. Are they schematically valid?

- Can we characterize the substitutions σ that are "problematic" in the sense that for some valid φ, $\sigma(\varphi)$ is not valid, or characterize the problematic substitution instances themselves?

As a special case, there has been much discussion of the problem of characterizing those substitutions σ such that $\sigma(p \to \langle p \rangle p)$ is not valid [17, 30, 18]. For models of a single fully introspective agent, the answer is that *all* such substitutions use Moorean formulas [30]. In the general case for any valid φ, there is an obvious necessary syntactic condition on problematic substitutions: they must use *epistemic* formulas somehow. The question is how exactly.

Although we have focused in this paper on the problem of finding "problematic" substitutions, a further question is whether we can give a *finite axiomatization* of the set of principles that are always safe from such substitutions—the substitution core. In a continuation of this work [29], we present a finite axiomatization of the substitution core of PAL in a system of Uniform Public Announcement Logic (UPAL).

Acknowledgements
We wish to thank Johan van Benthem for stimulating our interest in the topic of this article and the anonymous referees for helpful comments. This article is based on our earlier conference paper [28] presented at the 2011 LORI-III Workshop in Guangzhou, China. We are grateful to the workshop organizers and participants for valuable discussion.

References

1. Aqvist, L.: 1973, 'Modal Logic with Subjunctive Conditionals and Dispositional Predicates'. *Journal of Philosophical Logic* **2**, 1–76.
2. Balbiani, P., A. Baltag, H. van Ditmarsch, A. Herzig, T. Hoshi, and T. de Lima: 2008, ''Knowable' as 'known after an announcement''. *The Review of Symbolic Logic* **1**, 305–334.
3. Ballarin, R.: 2005, 'Validity and Necessity'. *Journal of Philosophical Logic* **34**, 275–303.
4. Baltag, A., L. 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 98).* Morgan Kaufmann, pp. 43–56.
5. van Benthem, J.: 2004, 'What One May Come to Know'. *Analysis* **64**(2), 95–105.
6. van Benthem, J.: 2006a, 'One is a Lonely Number: Logic and Communication'. In: Z. Chatzidakis, P. Koepke, and W. Pohlers (eds.): *Logic Colloquium '02.* ASL & A.K. Peters, pp. 96–129.
7. van Benthem, J.: 2006b, 'Open Problems in Logical Dynamics'. In: D. Gabbay, S. Goncharov, and M. Zakharyashev (eds.): *Mathematical Problems from Applied Logic I.* Springer, pp. 137–192.

8. van Benthem, J.: 2011, *Logical Dynamics of Information and Interaction*. Cambridge University Press.
9. van Benthem, J., J. van Eijck, and B. Kooi: 2006, 'Logics of communication and change'. *Information and Computation* **204**(11), 1620–1662.
10. Blackburn, P., M. de Rijke, and Y. Venema: 2001, *Modal Logic*. Cambridge University Press.
11. Buss, S. R.: 1990, 'The Modal Logic of Pure Provability'. *Notre Dame Journal of Formal Logic* **31**(2), 225–231.
12. Carnap, R.: 1946, 'Modalities and Quantification'. *The Journal of Symbolic Logic* **11**(2), 33–64.
13. Ciardelli, I. A.: 2009, 'Inquisitive Semantics and Intermediate Logics'. Master's thesis, University of Amsterdam. ILLC Master of Logic Thesis Series MoL-2009-11.
14. Davies, M. and L. Humberstone: 1980, 'Two Notions of Necessity'. *Philosophical Studies* **38**, 1–30.
15. del Cerro, L. F. and A. Herzig: 1996, 'Combining Classical and Intuitionistic Logic – or: intuitionistic implication as a conditional'. In: F. Baader and K. Schulz (eds.): *Frontiers of Combining Systems*. Kluwer Academic Publishers, pp. 93–102.
16. van Ditmarsch, H.: 2003, 'The Russian cards problem'. *Studia Logica* **75**, 31–62.
17. van Ditmarsch, H. and B. Kooi: 2006, 'The Secret of My Success'. *Synthese* **151**, 201–232.
18. van Ditmarsch, H., W. van der Hoek, and P. Iliev: 2011, 'Everything is Knowable – How to Get to Know *Whether* a Proposition is True'. *Theoria* **78**(2), 93–114.
19. van Ditmarsch, H., W. van der Hoek, and B. Kooi: 2008, *Dynamic Epistemic Logic*. Springer.
20. Fagin, R., J. Y. Halpern, Y. Moses, and M. Y. Vardi: 1995, *Reasoning about Knowledge*. MIT Press.
21. Fitch, F. B.: 1963, 'A Logical Analysis of Some Value Concepts'. *The Journal of Symbolic Logic* **28**(2), 135–142.
22. Gerbrandy, J. and W. Groenevelt: 1997, 'Reasoning about Information Change'. *Journal of Logic, Language and Information* **6**(2), 147–169.
23. Halpern, J. Y.: 1996, 'Should Knowledge Entail Belief?'. *Journal of Philosophical Logic* **25**, 483–494.
24. Hintikka, J.: 1962, *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. Cornell University Press.
25. Holliday, W. H.: 2012, 'Hintikka's Anti-Performatory Effect and Fitch's Paradox of Knowability'. Manuscript.
26. Holliday, W. H.: 2013a, 'Epistemic Closure and Epistemic Logic I: Relevant Alternatives and Subjunctivism'. *Journal of Philosophical Logic*. Forthcoming.
27. Holliday, W. H.: 2013b, 'Epistemic Logic and Epistemology'. In: S. O. Hansson and V. F. Hendricks (eds.): *Handbook of Formal Philosophy*. Springer. Forthcoming.
28. Holliday, W. H., T. Hoshi, and T. F. Icard, III: 2011, 'Schematic Validity in Dynamic Epistemic Logic: Decidability'. In: H. van Ditmarsch, J. Lang, and S. Ju (eds.): *Proceedings of the Third International Workshop on Logic, Rationality and Interaction (LORI-III)*, Vol. 6953 of *Lecture Notes in Artificial Intelligence*. Springer, pp. 87–96.

29. Holliday, W. H., T. Hoshi, and T. F. Icard, III: 2012, ‘A Uniform Logic of Information Dynamics’. In: T. Bolander, T. Braüner, S. Ghilardi, and L. Moss (eds.): *Advances in Modal Logic*, Vol. 9. College Publications, pp. 348–367.
30. 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*, Vol. 8. College Publications, pp. 178–199.
31. Hoshi, T. and A. Yap: 2009, ‘Dynamic epistemic logic with branching temporal structures’. *Synthese* **169**(2), 259–281.
32. Mascarenhas, S.: 2009, ‘Inquisitive Semantics and Logic’. Master’s thesis, University of Amsterdam. ILLC Master of Logic Thesis Series MoL-2009-18.
33. Meyer, J.-J. Ch. and W. van der Hoek: 1995, *Epistemic Logic for AI and Computer Science*, Vol. 41 of *Cambridge Tracts in Theoretical Computer Science*. Cambridge University Press.
34. Moore, G. E.: 1942, ‘A Reply to My Critics’. In: P. A. Schilpp (ed.): *The Philosophy of G.E. Moore*. Northwestern University, pp. 535–677.
35. Plaza, J.: 1989, ‘Logics of Public Communications’. In: M. Emrich, M. Pfeifer, M. Hadzikadic, and Z. Ras (eds.): *Proceedings of the 4th International Symposium on Methodologies for Intelligent Systems*. Oak Ridge National Laboratory, pp. 201–216.
36. Quine, W. V. O.: 1960, *Word and Object*. MIT Press.
37. Reynolds, M.: 2001, ‘An Axiomatization of Full Computation Tree Logic’. *The Journal of Symbolic Logic* **66**(3), 1011–1057.
38. Schurz, G.: 2005, ‘Logic, matter of form, and closure under substitution’. In: M. Bilkova and L. Behounek (eds.): *The Logica Yearbook*. Filosofia, pp. 33–46.
39. Segerberg, K.: 1973, ‘Two-Dimensional Modal Logic’. *Journal of Philosophical Logic* **2**(1), 77–96.
40. Wang, Y.: 2011, ‘On Axiomatizations of PAL’. In: H. van Ditmarsch, J. Lang, and S. Ju (eds.): *Proceedings of the Third International Workshop on Logic, Rationality and Interaction (LORI-III)*, Vol. 6953 of *Lecture Notes in Artificial Intelligence*. Springer, pp. 314–327.
41. Williamson, T.: 2000, *Knowledge and its Limits*. Oxford University Press.