Cultural knowledge evolution in dynamic epistemic logic

Line van den Berg

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ GRENOBLE ALPES

Spécialité : Mathématiques et Informatique

Arrêté ministériel : 25 mai 2016

Présentée par

Line VAN DEN BERG

Thèse dirigée par **Jérôme EUZENAT**,DIRECTEUR DE RECHERCHE, Université Grenoble Alpes
et co-encadrée par **Manuel ATENCIA ARCAS**, UGA

préparée au sein du **Laboratoire d'Informatique de Grenoble**
dans **l'École Doctorale Mathématiques, Sciences et technologies de l'information, Informatique**

L'évolution culturelle de la connaissance en logique épistémique dynamique

Cultural knowledge evolution in dynamic epistemic logic

Thèse soutenue publiquement le **29 octobre 2021**,
devant le jury composé de :

Monsieur JERÔME EUZENAT
DIRECTEUR DE RECHERCHE, INRIA CENTRE GRENOBLE-RHONE-ALPES, Directeur de thèse
Monsieur ANDREAS HERZIG
DIRECTEUR DE RECHERCHE, CNRS DELEGATION OCCITANIE OUEST, Rapporteur
Monsieur MARCO SCHORLEMMER
CHERCHEUR HDR, IIIA-CSIC, Rapporteur
Madame SOPHIE PINCHINAT
PROFESSEUR DES UNIVERSITES, UNIVERSITE RENNES 1, Examinatrice
Madame SONJA SMETS
PROFESSEUR, Universiteit van Amsterdam, Examinatrice

Monsieur PIERRE GENEVES
DIRECTEUR DE RECHERCHE, CNRS DELEGATION ALPES, Président

Cultural Knowledge Evolution in Dynamic Epistemic Logic

Line van den Berg

Voilà, voilà, voilà, voilà qui je suis

BARBARA PRAVI

Acknowledgements

There are a lot of similarities between writing a PhD and going on a climbing trip. In advance, you are full of ambitious plans and picture the perfect route to the summit: this crazy-looking north face, this insane ridge. But when you arrive at the basecamp, new elements are added to the equation, like conditions, weather, acclimatization, fuelling the body with enough nutrients, etc. At times, conditions are bad, or it is raining for days, and you'd be tempted to give up. Adaptation and patience are key. One day, the sky will clear and you will find a way to the top, even if it is not exactly the mountain or route you envisioned. It will be a fight, but setting foot on the summit vanishes the doubts, the struggles, and happiness prevails. Voilà, this PhD is the route I climbed to "Mont Docteur".

Like in climbing, success in doing a PhD is very dependent on external circumstances, most notably the people around you. First of all, I would like to thank my supervisors Jérôme and Manuel. They have supported me through the highs and lows of writing this thesis. I am really grateful for their trust and space to develop my own ideas and bring logic to cultural knowledge evolution, while also keeping me sharp and teach me how to count (I will use n-tuples from now on). Finally, thank you for teaching me papers are never finished, just deadlines are. I would also like to thank the complete mOeX team for their support, especially in the lockdowns when we continued our meetings online. Finally, I am grateful to MIAI Grenoble Alpes for financing part of my PhD.

I am also happy to have the following people on my committee: Andreas Herzig, who already gave me helpful feedback halfway through this thesis as a part of my thesis follow-up committee, Marco Schorlemmer, whose work on interaction-situated alignments has always been useful to explain to friends and family what I am doing, Sonja Smets, who was my mentor at the Master of Logic and has been very helpful in finding my path in logic, Sophie Pinchinat, and Pierre Genevès.

During the thesis, I developed a game called *Class?*[1] with Jérôme Euzenat

[1]See `https://moex.inria.fr/mediation/class/index.html` to download the game.

to teach the concepts of cultural knowledge evolution to the broader public. I would like to thank everyone who came to play the game and, through playing, helped us to evolve it.

Then, I would like to thank the people who have played important roles in my life as a logician. Thank you Dora for lighting the flame of logic in me, and Dora and Jan for encouraging me to pursue a PhD. Furthermore, I want to thank Malvin to have continued gossiping[2] when I exchanged Amsterdam for the mountains. I hope we will share more talks about logic, academic life and biking in the future.

I moved to France for the mountains, I would like to stay for the people (except that I found a postdoc in Bern). *Un grand merci* to the Weird Injured Climbers Group[3], aka the *Dommage pas de Fromage* group, aka the group who cannot agree on a name and therefore changes it as many times as the French say "bah-uh-oui": Camille, Amber, Michael, Nathan, Jacques, Louise, Felix, Jonathan and Isabelle – you guys are truly, truly amazing. I cannot thank you enough for all the laughs, love, climbing and craziness we shared, and for cooking me dinner when all basic life tasks got swallowed by the thesis writing process. I hope, and am sure, we will share many more great adventures in life.

Paul, thank you for all the support in the past three years. Your curls are enough to make me smile for days. Max, thanks for keeping it real: it's just left, right, left, right, and you'll eventually reach the summit. (Ok, maybe sometimes it's twice left...) I am psyched for more pumping drytooling sessions. Laila, it is just great to have another Dutchie around, especially you. Alex and Clément, I am very happy to have had you as flatmates, you are missed. Jean-Marc, thank you and your family for being my little *famille française*. Fay, I am so stoked to have met my alpine twin and am so excited for new adventures!

Then, a special shout out to my flatmates Joseba and Louisa. Joseba, your hugs make every bad day into a good one, no exceptions. Luisa, you always find a positive note in everything. Thank you both for all the spontaneous bike rides, talks, games and beers.

I would not have survived the French culture without sometimes being reminded of the Dutch sense of humor, directness and efficiency (sorry, but it's true!). Thanks to all my Dutch friends who kept visiting to climb and spend time together, especially Jorian and Mats. We don't call ourselves 'team super' for no reason, it is really super to have you in my life. Jorian,

[2]This is a little joke concerning a paper [21] we jointly wrote on the topic of dynamic gossip.

[3]It is quite ironic that I managed to be injured at my own defence.

Mats, Bas, Nicole and Danny, unfortunately our perseverance to plan an expedition did not pay off, but I am sure we will get our revanche in the future one way or another. Martin, you have become one of my dearest friends, thank you for always being there for me. Lastly, thanks to all other rope partners (as well as in climbing as in life). The bond that is developed in the mountains is a very special one.

Finally, I would never be here without my family: Trudy, Albert, Joran and Remo. Dear Trudy and Albert, without you this thesis (nor I for that sake) would not exist. I am very proud to carry your genes, although at times the mixture can be a bit challenging (I am sure you know). You have always supported me in all my choices and followed me on all my crazy ideas, even if that meant being worried when I was sleeping on a tiny ledge high up in the mountains. Thanks for visiting me in Grenoble to conquer the cols and taste the French life, you are the greatest parents. Dear Joran and Remo, I am so proud to see what persons you have grown to. You both inspire me in many ways. Also, a great thank you also to the rest of my family, especially the Bergfreunde and Vosco's.

Before heading to the scientific part of this thesis, let me give one advice to those of you not so familiar with logic or computer science, to quote Diggy Dex (La vie est belle):

Je n'y comprends rien
Je vais rire jusqu'à la fin
La vie est belle

Last but not least, thanks to everyone who started reading my thesis and stopped here. Thanks to everyone who found this funny.

Contents

Chapter 1

Introduction

Agents are the subject of studying autonomous behavior. They may describe, reason and communicate about the world around them in different ways. Just as humans may speak different languages or, even when they do speak the same language, may use different words to denote the same thing ('soccer' versus 'football'), agents may use different vocabularies or assign different meanings to the terms they use. Both of these, the vocabularies and the meaning assigned, can be described by the *knowledge representations* of agents. A typical structure of these knowledge representations are *ontologies*, formal theories describing classes and objects and relations between them [93]. These relations denote which objects belong to which classes and how classes relate to each other.

Example 1.1. In Figure 1.1 an ontology is depicted for sports. The relation $\sqsubseteq$ denotes subsumption: $A \sqsubseteq B$ means that every A is B. The relation $\oplus$ denotes disjointness: $A \oplus B$ means that A and B have no overlap.

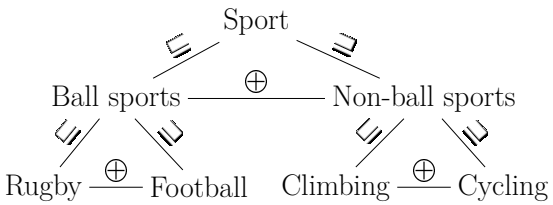

Figure 1.1: An ontology for sports, to which objects such as athletes or matches can be assigned.

When agents use different ontologies to represent what they know, an issue arises when they try to communicate. How do they understand each other if they do not use the same terms or, when they do, assign a different

meaning to them? In an analogy to humans, two persons that do not speak the same language will try to find other ways to interact: they use examples, body-language and learn from their mistakes. Even when they do speak the same language, it is widely known that 'soccer' and 'football' denote the same thing, not to be confused with the similar notion '(American) football' with a different meaning depending on geographical information.

Through experience, humans learn how to *translate* their own knowledge to the knowledge of others. Such translations are essential to achieve successful communication while preserving their differences, i.e. the heterogeneity of their knowledge. Of course, alternative to using these translations, humans could be required to use a single, common language to which a unique meaning is assigned. Pride and prejudice aside, this is not feasible nor desirable: some parts of the culture, in which language is rooted, may get lost and it comes at the price of autonomy. The same arguments apply to situations in which a single ontology cannot be enforced. To preserve heterogeneity, agents are then required, just as humans, to find a sort of "translation" between their ontologies. This is also called an *alignment* [50]: a set of correspondences between the terms of two ontologies.

Ontology matching algorithms have been developed to compute alignments and provide them to the agents [50]. However, they may output only partially correct or incomplete alignments, and there may be situations in which agents constantly evolve their ontologies [4]. For example, they may learn new terms from other agents or encounter them in their environment, or they may adapt the meaning they assign to these terms. As a consequence, alignments may become incorrect and incomplete. Therefore, with the evolution of their own ontologies, agents are required to evolve their alignments accordingly.

Example 1.2. In Figure 1.2 two ontologies are depicted for sports with an alignment between them. The relation $\sqsubseteq$ again denotes subsumption and the relation $\equiv$ is an abbreviation for $\sqsubseteq$ and $\sqsupseteq$. Notice that the alignment here is correct but not complete when we consider the meaning assigned to the terms in the English language.

Experimental Cultural Evolution

Experimental cultural evolution aims at studying the mechanisms that agents use to evolve their culture in situated environments: environments in which agents simultaneously use their culture to accomplish a (joint) task, called a game, and evolve it when failures occur [5]. It takes inspiration from the

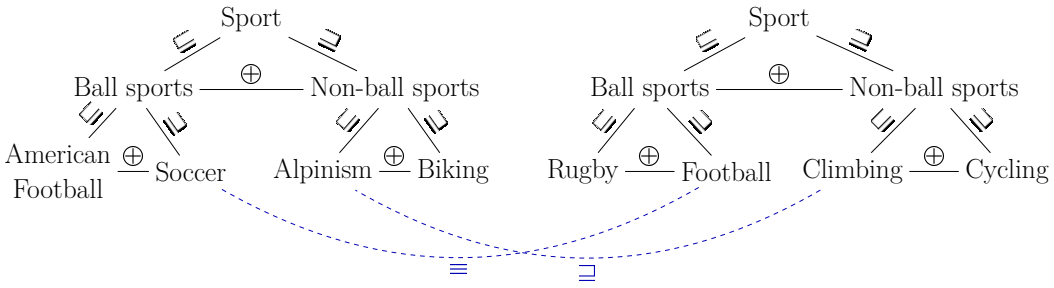

Figure 1.2: Two ontologies for sports and an alignment between them.

theory of evolution, which studies populations subjected to natural selection leading to the discovery and fixation of fitter variants [90], and applies it to culture, where culture is taken as an intellectual artefact shared among such a population. Experiments in this domain typically observe a group of agents that evolve their culture through a predetermined protocol. The goal of such experiments is to discover to what common state the agents converge and which properties hold at this state.

The ideas of experimental evolution have been successfully applied to language [91], showing how language, which may be seen as a culture, can be shared and evolved in a society through communication. Beyond language, their contribution is a precise experimental methodology: a group of agents play, at random, an "interaction" game and the outcome of the game determines whether agents adapt their language. Through monitoring the state of the system, for example what the success rate is of the game, convergence to a stable state can be established experimentally.

Experimental cultural evolution can also be applied to knowledge representations to obtain a plausible model of knowledge transmission. The Alignment Repair Game (ARG) [46] has been introduced for this purpose. ARG applies the methodology of experimental cultural evolution to knowledge. In particular to the knowledge of agents of how to translate the terms in their ontologies to terms in other agents' ontologies, i.e. to their alignments. In ARG, agents, with different ontologies, communicate and, in parallel, evolve the alignments between their ontologies through local corrective actions whenever communication fails. This means that agents learn on the fly and repair mistakes when they occur. The local corrective actions that agents perform are called *adaptation operators*. Adaptation operators specify, given the failure of a certain correspondence, what the agents should do. The adaptation operators discussed in [46, 49] have in common that they discard the failing correspondence and typically, they provide an improved

correspondence for agents to adopt. Hence, they are a strategy for agents how to evolve their alignments.

Through experiments with the different adaptation operators for ARG, their properties can be assessed and they can be compared. Overall, the experiments showed that agents converge towards successful communication and improve their alignments [46, 49]. This means that fewer and fewer failures occur and a stable state is reached in which all communication taking place between agents is successful. In this state, the alignments do not necessarily correspond to the *reference alignment*, the "perfect" alignment between their ontologies, but is sufficient for agents to communicate successfully. ARG has been extended to allow agents to craft alignments from scratch via introducing random correspondences [48] and to reach fully correct correspondences [49].

A Logical Model for ARG

Experimental cultural knowledge evolution helps to understand how agents evolve their knowledge: through simulations with a large number of games or rounds, properties such as convergence can be studied and established experimentally. However, such simulations and experiments are not sufficient to understand the formal, logical properties of cultural knowledge evolution, whether for example the adaptation operators applied in ARG are formally correct, complete or redundant.

This thesis introduces a logical model for ARG that can examine the formal properties of the adaptation operators in ARG. We define Dynamic Epistemic Ontology Logic (DEOL), which combines the dynamic (announcements, radical and conservative upgrades) and epistemic (the definitions of knowledge and belief) aspects of Dynamic Epistemic Logic with classifications and class relations of a simple Description Logic, and a faithful translation from ARG to DEOL. This translation (a) encodes ARG ontologies, (b) maps agents' ontologies and alignments to knowledge and beliefs and (c) captures the adaptation operators through announcements and conservative upgrades. With this translation, we investigate how the adaptation operators applied by ARG agents when a failure occurs compare to the mechanisms in logic for agents to evolve their knowledge and beliefs. This gives rise to formal definitions of correctness, completeness and redundancy. We prove that all but one adaptation operator are correct, all adaptation operators are incomplete and some adaptation operators are partially redundant. With these results, this thesis bridges a very practical implementation of *adaptive agents*, used in simulations like for example in ARG, with a dynamic

epistemic model of these agents, also referred to as *logical agents*.

Awareness

Despite the lack of formal properties, experiments have shown that ARG works quite well in practice, suggesting that agents do not need to be logical to communicate successful. There are therefore two correct and compatible ways to interpret the results: either the ARG agents use a sub-logical behavior to evolve their alignments, or the logical model of ARG in DEOL is insufficient to describe their behavior. Implementing agents in ARG that reason more closely to the logic may improve their behavior, yet, this was not the goal of the initial experiments nor of this thesis, which follows the second interpretation.

We identify three differences between the adaptive agents and logical agents: (1) the adaptive agents reason locally while the logical agents reason globally, (2) logical agents share a fixed vocabulary, preventing them from using heterogeneous knowledge representations like adaptive agents, and (3) the adaptive agents are unable to remember individual cases because they focus on general knowledge, whereas logical agents cannot discard these.

In order to address these differences, we introduce awareness, based on partial valuation functions and weakly reflexive relations, called Partial Dynamic Epistemic Logic (ParDEL), which allows to drop the assumption of shared, fixed vocabularies. In ParDEL, propositions can be true, false or *undefined* and can therefore be used to model agents that use different vocabularies to model their knowledge and beliefs. As such, semantic heterogeneity between agents can be preserved, while successful communication can be achieved through raising awareness modalities. Furthermore, awareness enables agents to discard evidence in favour of general knowledge via forgetting modalities.

Based on this notion of awareness, we define an alternative translation of ARG under which we prove that the adaptation operators are correct, complete for ARG states consisting of two agents and no longer (partially) redundant, therefore confirming that DEOL is insufficient to model cultural knowledge evolution. Furthermore, the notion of awareness is used to show that, to communicate successfully in ARG, agents do not need full awareness of the vocabularies used by other agents.

Contribution

The contribution of this thesis is two-fold. The Alignment Repair Game is modeled and evaluated using logic, establishing the formal properties of the adaptation operators. Furthermore, an independent model of awareness is introduced on which raising awareness and forgetting modalities are defined. Together, they pave the way for defining a theoretical model of cultural knowledge evolution.

Outline

Chapter 2: Basics

This thesis uses ideas from Dynamic Epistemic Logic (DEL) to model the Alignment Repair Game (ARG) and explore its formal properties. In this chapter the basic building blocks are presented and the position of the thesis with respect to the related work is discussed.

Chapter 3: A Logical Model for the Alignment Repair Game

The Alignment Repair Game (ARG) is translated to Dynamic Epistemic Ontology Logic (DEOL), a variant of Dynamic Epistemic Logic that uses classifications and class relations of a simple Description Logic language as propositions. Then, this translation is used to define and establish the formal properties of the adaptation operators in ARG.

Chapter 4: Fundamental Differences between Adaptive Agents and Logical Agents

Three differences between the adaptive agents (those playing ARG) and logical agents (those in the logical model of ARG in DEOL) are identified and discussed: (1) the adaptive agents reason locally while logical agents reason globally, (2) logical agents share a fixed vocabulary, preventing them from using heterogeneous knowledge representations like adaptive agents, and (3) the adaptive agents are unable to remember individual cases because they focus on general knowledge, whereas logical agents cannot discard these.

Chapter 5: Agent Awareness

In this chapter, we introduce a novel framework of awareness for Dynamic Epistemic Logic, called Partial Dynamic Epistemic Logic (ParDEL), in order to overcome one of the differences concerning the vocabularies used by agents. We define and formalize the properties of awareness and introduce a modality for raising awareness.

Chapter 6: Forgetting

We use the notion of awareness to overcome another difference between adaptive agents and logical agents and introduce two forgetting modalities for ParDEL: forgetting awareness and forgetting truth.

Chapter 7: Formal Properties of the Adaptation Operators Revisited

We define an alternative translation of ARG incorporating awareness, raising awareness and forgetting, and re-examine the formal properties of the adaptation operators with respect to this translation. Then, how awareness evolves on ARG will be studied and linked to successful communication.

Chapter 8: Conclusion

In this chapter, we present the perspectives of this work for ARG as well as beyond ARG.

Origin of Materials

Chapter 3 is based on two papers, where the latter is an extended version of the former:

- [17] Van den Berg, L., Atencia, M., Euzenat, J.: Agent ontology alignment repair through dynamic epistemic logic. In: Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems, pp. 1422-1430. International Foundation for Autonomous Agents and Multiagent Systems (2020)
- [19] Van den Berg, L., Atencia, M., Euzenat, J.: A logical model for the ontology alignment repair game. Autonomous Agents and Multi-Agent Systems **35**(2), 1-34 (2021)

Chapter 5 is based on two papers, where the latter is an extended version of the former:

- [18] Van den Berg, L., Atencia, M., Euzenat, J.: Unwareness in multi-agent systems with partial valuations. In: 10th AAMAS workshop on Logical Aspects of Multi-Agent Systems (LAMAS). No commercial editor. (2020)
- [20] Van den Berg, L., Atencia, M., Euzenat, J.: Raising awarness without disclosing truth (submitted)

Chapter 6 is based on:

- [16] Van den Berg, L.: Forgetting agent awareness: a partial semantics approach. In: 4th Women in Logic workshop (WiL), pp. 18-21. No commercial editor. (2020)

Chapter 2

Basics

This thesis uses ideas from Dynamic Epistemic Logic (DEL) to model a multi-agent game represented by ontologies and alignments. In this chapter the basic building blocks for that purpose are presented.

First, ontologies and alignments are defined, and the Alignment Repair Game (ARG) is introduced. Then, we move to DEL that forms the basis for the translation of ARG into logic and define models, satisfiability and event models for dynamic upgrades. After that, we will discuss some other approaches to using logic to model multi-agent systems, or games. Finally, the related work on awareness will be introduced.

2.1 Ontologies and Alignments

In order to accomplish their tasks, agents often maintain a representation of the world they live in. Using an ontology for that purpose is commonplace [50]. Ontologies are a tool to represent information about entities through classifications that categorize the entities in classes and relate them.

An ontology is based on a signature.

Definition 2.1 (Ontology signature)**.** An *ontology signature*, or *signature*, $sig(\mathcal{O})$ is a pair $\langle \mathcal{C}, \mathcal{D} \rangle$ such that $\mathcal{C}$ is a set of class names, with $\top \in \mathcal{C}$, and $\mathcal{D}$ is a set of object names.

We use uppercase letters (C, D) to denote elements of $\mathcal{C}$ and lowercase letters (o, o') to denote elements of $\mathcal{D}$. *Statements* or *formulas* relate elements of a signature through relations of subsumption ($C \sqsubseteq D$ or $C \sqsupseteq D$), disjointness ($C \oplus D$) or membership ($C(o)$). We may say that C is equivalent to D ($C \equiv D$) to abbreviate $C \sqsubseteq D$ and $D \sqsubseteq C$.

An ontology, denoted by $\mathcal{O}$, over a signature, is a set of such statements constraining the interpretation of the objects and classes of the signature.

Formally, the ontology can be expressed as a knowledge base in Description Logics [6]. In this thesis, we restrict ourselves to the very simple ontologies manipulated within the Alignment Repair Game (ARG). They express the minimum necessary to allow agents to play the game. They do not consider roles, nor complex classes, and they are organised into dichotomic trees which may be informally described as follows, given an ontology signature $\langle \mathcal{C}, \mathcal{D} \rangle$:

- Each class in $\mathcal{C}$ is assigned a node in a binary tree rooted at $\top$, whenever a class C is a child of a class D, then $C \sqsubseteq D \in \mathcal{O}$,
- for every $C, D \in \mathcal{C}$, if C and D are siblings then $C \oplus D \in \mathcal{O}$,
- for every leaf C of the tree, there exists $o \in \mathcal{D}$ such that $C(o) \in \mathcal{O}$, and
- each $o \in \mathcal{D}$, is attached to a most specific node C, i.e. $C(o) \in \mathcal{O}$.

This is illustrated by Figure 2.1. We will consider several agents, named a, b, c, etc., each having their own ontology. We will use subscripts to identify the agents for the associated concepts: $\mathcal{O}_a$, $\mathcal{C}_a$, $mgcx_a(C, o)$, etc.

ARG ontologies are formally defined in Definition 2.2 on which proofs are based. Such constraints may be obtained in other ways, but the present ones are sufficient for this thesis.

Definition 2.2 (ARG ontology)**.** An *ARG ontology* $\mathcal{O}$ over a signature $\langle \mathcal{C}, \mathcal{D} \rangle$ is a finite set of axioms:

$$C \sqsubseteq D \mid C \oplus D \mid C(o)$$

with $C, D \in \mathcal{C}$ different class names and $o \in \mathcal{D}$ an object name, such that:

1. $\forall C \in \mathcal{C}$: $\top \sqsubseteq C \notin \mathcal{O}$,
2. $\forall C \in \mathcal{C} \backslash \{\top\}$ $\exists! D \in \mathcal{C}$: $C \sqsubseteq D \in \mathcal{O}$,
3. $\forall D \in \mathcal{C}$, either one of the following holds:
   (a) $\exists! \langle C, C' \rangle \in \mathcal{C} \times \mathcal{C}$ with $C \sqsubseteq D \in \mathcal{O}$, $C' \sqsubseteq D \in \mathcal{O}$ and $C \oplus C' \in \mathcal{O}$, or
   (b) $\nexists C \in \mathcal{C}$ such that $C \sqsubseteq D \in \mathcal{O}$, and $\exists o \in \mathcal{D}$; $D(o) \in \mathcal{O}$.
4. $\forall o \in \mathcal{D}$, $\exists! C \in \mathcal{C}$; $C(o) \in \mathcal{O}$;
5. $\forall C \in \mathcal{C}$, $\nexists C_0, \ldots C_n$; $\forall i \in [1, n]$, $C_i \sqsubseteq C_{i-1} \in \mathcal{O}$ and $C_0 = C = C_n$;
6. $\mathcal{O}$ contains no other axiom.

The constraints in Definition 2.2 specify that (1) there is a class assigned no superclass, (2) each class, but $\top$, is assigned to be subsumed by another class, (3) all classes are either (a) assigned to subsume a pair of classes, that are assigned to be disjoint, or (b) they are not assigned to subsume any class but an object is assigned to belong to them, (4) each object has a unique most specific class to which it is assigned to belong, and (5) there is no cycle in the subsumption assertions. Notice that in 3(b), the 'and' is the main connective, i.e. the intended meaning, using brackets, is: ($\nexists C \in \mathcal{C}$ such that $C \sqsubseteq D \in \mathcal{O}$) and ($\exists o \in \mathcal{D}$; $D(o) \in \mathcal{O}$).

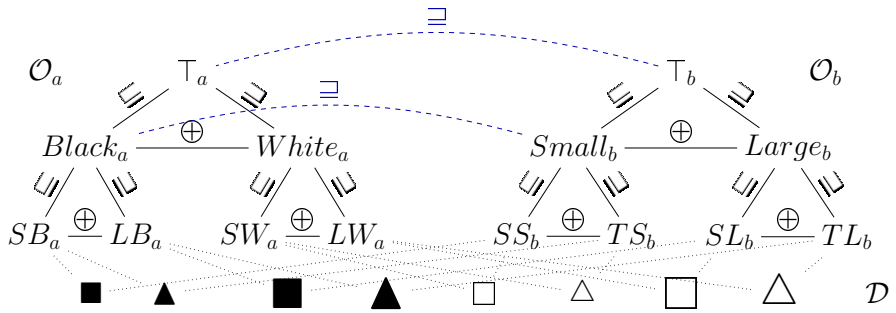

Figure 2.1: Two ARG ontologies, $\mathcal{O}_a$ on the left and $\mathcal{O}_b$ on the right, and an alignment A_{ab} (in dashed blue) between them. Membership between objects and classes are rendered by dotted edges. Relationships between classes are rendered by solid edges. The class names of the leaf classes have an intended use for the agents that use them: SB_a has the intended meaning of being the class, in agent a's ontology, with all objects that are both small and black; LW_a all objects that are large and white; SS_b is the class, in agent b's ontology, with all objects that are small and squared; TL_b all objects that are triangular and large; etc. Of course, these intended uses are not accessible for other agents.

In the following, the word 'ontology' denotes an ARG ontology. For a given ontology $\mathcal{O}$ over a signature $\langle \mathcal{C}, \mathcal{D} \rangle$, we will also write $C \in \mathcal{O}$ for $C \in \mathcal{C}$ when $\mathcal{C}$ is clear from the context. Similarly we write $o \in \mathcal{O}$ for $o \in \mathcal{D}$ when $\mathcal{D}$ is clear from the context.

Definition 2.2 constrains the syntax of ARG ontologies. An *interpretation* is provided to give meaning to the elements of the signature.

Definition 2.3 (Ontology Interpretation)**.** Given a signature $\langle \mathcal{C}, \mathcal{D} \rangle$, an *ontology interpretation* I is a tuple $I = \langle \Delta, \cdot^I \rangle$ such that Δ is a non-empty *domain*, a set of objects, and $\cdot^I$ is a function assigning to object names $o \in \mathcal{D}$ an element of the domain Δ ($\cdot^I : \mathcal{D} \to \Delta$), and to class names $C \in \mathcal{C}$ a subset of Δ ($\cdot^I : \mathcal{C} \to \mathcal{P}(\Delta)$), such that $\top^I = \Delta$.

Definition 2.4 (Formula Satisfaction)**.** Let $\mathcal{O}$ be an ontology and let I be an interpretation over the signature of $\mathcal{O}$, satisfaction ($\models$) with respect to I is defined as follows:

$$I \models C \sqsubseteq D \text{ iff } C^I \subseteq D^I$$
$$I \models C \oplus D \text{ iff } C^I \cap D^I = \varnothing$$
$$I \models C(o) \text{ iff } o^I \in C^I$$

Furthermore, we say that a formula ϕ is *satisfiable* on $\mathcal{O}$, written $\mathcal{O} \models \phi$, if there exists an interpretation I over the signature of $\mathcal{O}$ that satisfies ϕ.

We may use $I \models C \sqsubset D$ whenever $I \models C \sqsubseteq D$ but $I \not\models C \sqsupseteq D$, $I \models C \sqsharp D$ whenever $I \not\models C \sqsubseteq D$ and $I \models C \between D$ whenever $I \not\models C \oplus D$ and say that C and D are overlapping.

As usual, an interpretation satisfying all the axioms of an ontology is called a *model* of that ontology; an ontology for which there does not exist a model is *inconsistent*; and an ontology $\mathcal{O}$ entails a statement ϕ if all models of the ontology satisfy this statement (noted $\mathcal{O} \models \phi$).

In the case of ARG, ontologies always have models. This is a good reason for ARG agents to never change their ontologies. Given an ARG ontology $\mathcal{O}$, we define the standard interpretation of $\mathcal{O}$ as follows.

Definition 2.5 (Standard Model)**.** Let $\mathcal{O}$ be an ARG ontology over a signature $\langle \mathcal{C}, \mathcal{D} \rangle$. Let $\hat{I}_0$ be the interpretation defined by

1. $\Delta^{\hat{I}_0} = \mathcal{D}$
2. $o^{\hat{I}_0} = o$ for every $o \in \mathcal{D}$
3. $C^{\hat{I}_0} = \{o \in \mathcal{D} : C(o) \in \mathcal{O}\}$ for every $C \in \mathcal{C}$

Given $\hat{I}_k$ ($k \geqslant 0$), we define $\hat{I}_{k+1}$ as an extension of $\hat{I}_k$ by applying the following rule:

$$\text{if } C \sqsubseteq D \in \mathcal{O} \text{ and } o \in \mathcal{D} \text{ s.t. } o \in C^{\hat{I}_k} \text{ and } o \notin D^{\hat{I}_k} \text{ then } o \in D^{\hat{I}_{k+1}} \tag{2.1}$$

Then the standard model $\hat{I}$ is defined as $\hat{I}_n$ such that $\hat{I}_n = \hat{I}_m$ for every $m \geqslant n$.

In the following, we will use the term *standard interpretation* for classes C to denote the interpretation of C in the standard model.

Corollary 2.1 ([19])**.** Every ARG ontology is consistent.

We introduce some notation useful for defining the Alignment Repair Game (ARG).

Definition 2.6. For any ARG ontology $\mathcal{O}$ of signature $\langle \mathcal{C}, \mathcal{D} \rangle$:

(a) For each class $C \in \mathcal{C} \backslash \{\top\}$, the *most specific superclass of* C is the class $D \in \mathcal{C}$ defined by:

$$\mathcal{O} \models C \sqsubset D \text{ and } \forall C' \in \mathcal{C} : \ \mathcal{O} \models C \sqsubset C' \Rightarrow \mathcal{O} \models D \sqsubseteq C'$$

It is denoted by $msc(C)$.

(b) For each object $o \in \mathcal{D}$, the *most specific class compatible with* o is the class $C \in \mathcal{O}$ defined by:

$$\mathcal{O} \models C(o) \text{ and } \forall C' \in \mathcal{C} : \mathcal{O} \models C'(o) \Rightarrow \mathcal{O} \models C \sqsubseteq C'$$

It is denoted by $msc(o)$.

(c) For each class $C \in \mathcal{C} \backslash \{\top\}$ and for each object $o \in \mathcal{D}$, the *most specific superclass of* C *compatible with* o is $D \in \mathcal{C}$ defined by:

$$\mathcal{O} \models C \sqsubset D \text{ and } \mathcal{O} \models D(o) \text{ and } \forall C' \in \mathcal{C} : \mathcal{O} \models C \sqsubset C' \wedge \ \mathcal{O} \models C'(o)$$

$$\Rightarrow \mathcal{O} \models D \sqsubseteq C'$$

It is denoted by $mscc(C, o)$.

(d) For each class $C \in \mathcal{C}$ and each object $o \in \mathcal{D}$, the *set of most general subclasses of* C *incompatible with* o, is the set defined by:

$$\{D \in \mathcal{C} \mid \mathcal{O} \models D \sqsubseteq C, \ \mathcal{O} \not\models D(o), \ \forall C' \in \mathcal{C} : (\mathcal{O} \models C' \sqsubseteq C \wedge \mathcal{O} \not\models C'(o))$$

$$\Rightarrow \mathcal{O} \models C' \sqsubseteq D\}$$

It is denoted by $mgcx(C, o)$.

Let us consider an example.

Example 2.1. Considering the ontologies of Figure 2.1, it can be observed that:

- $msc(SW_a) = White_a$: the most specific superclass of SW_a;
- $msc(\triangle) = SW_a$: the most specific class of $\triangle$;
- $mscc(Black_a, \triangle) = \top_a$: the most specific superclass of $Black_a$ compatible with $\triangle$, and

- $mgcx(Small_b, \triangle) = \{SS_b\}$: the most general subclasses of $Small_b$ incompatible with $\triangle$.

The concepts in Definition 2.6 are well-defined for ARG ontologies.

Lemma 2.1 ([19])**.** For any ARG ontology $\mathcal{O}$ of signature $\langle \mathcal{C}, \mathcal{D} \rangle$:

(a) For all classes $C \in \mathcal{C} \backslash \{\top\}$, $msc(C)$ exists and is unique;

(b) For all objects $o \in \mathcal{D}$, $msc(o)$ exists and is unique;

(c) For all classes $C \in \mathcal{C} \backslash \{\top\}$ and all objects $o \in \mathcal{D}$, $mscc(C, o)$ exists and is unique;

(d) For all classes $C \in \mathcal{C}$ and all objects $o \in \mathcal{D}$, $mgcx(C, o)$ exists and is unique.

Ontologies are used for many applications, such as database integration, semantic web services, social networks and e-commerce [53]. They are often considered as a complete solution for knowledge sharing between agents, but this only holds when agents have complete knowledge about the ontologies used by others. Imposing a single universally shared common ontology would be a straightforward way to establish this. However, autonomous agents, developed from different sources or learning autonomously, will naturally adopt different ontologies. Therefore, such an approach would require all involved parties to reach an agreement on the ontology to use, and discard their own ontologies. This comes at the price of autonomy, heterogeneity, diversity and privacy [28]. It is thus reasonable to consider that not all ontologies have to be shared by agents.

The heterogeneity of ontologies can be a problem, in particular when agents need to communicate about a common environment [93]: how do they understand each other if they express their knowledge in different ways? Or even if they use the same terms, they might misunderstand each other when using these terms to assign a different meaning to them. To preserve heterogeneity of agents' ontologies, a common method to ensure agent intelligibility when communicating relies on ontology alignments [50, 89]. Ontology alignments, or in short alignments, express relations between concepts and relations occurring in different ontologies. They are tools that allow agents to translate their knowledge with respect to the ontology of other agents. Therefore, alignments can be used by agents to interpret other agents' messages and therefore establish successful communication.

Formally, alignments are sets of correspondences between classes of two agents' ontologies.

Definition 2.7 (Ontology Alignment)**.** An *alignment* $A_{ab} \subseteq \mathcal{C}_a \times \mathcal{C}_b \times \{\sqsubseteq, \sqsupseteq\}$ between two ontologies $\mathcal{O}_a$ and $\mathcal{O}_b$ over signatures $\langle \mathcal{C}_a, \mathcal{D}_a \rangle$ and $\langle \mathcal{C}_b, \mathcal{D}_b \rangle$ is a set of correspondences $\langle C_a, C_b, R \rangle$ where C_a and C_b are class names belonging to $\mathcal{C}_a$ and $\mathcal{C}_b$, respectively, and $R \in \{\sqsubseteq, \sqsupseteq\}$.

We also write $C_a R C_b \in A_{ab}$ for $\langle C_a, C_b, R \rangle \in A_{ab}$. In this thesis, alignments are shared between the involved agents and we consider that class names are all disjoint ($\mathcal{C}_a \cap \mathcal{C}_b = \varnothing$) but object names are the same ($\mathcal{D}_a = \mathcal{D}_b = \mathcal{D}$).

The semantics for alignments used here is called the reduced semantics [50]. This semantics selects the pairs of models of each ontologies that satisfy the alignments.

However, in ARG, an agent is only aware of the constraints on her ontology. She will thus interpret alignments only with respect to her ontology. For instance, if $\langle C, D, \sqsupseteq \rangle \in A_{ab}$, $C \sqsubseteq C' \in \mathcal{O}_a$ and $D' \sqsubseteq D \in \mathcal{O}_b$, then A_{ab} and both ontologies $\mathcal{O}_a$ and $\mathcal{O}_b$ entail $C' \sqsupseteq D'$. However, the agents cannot access the other agents' ontologies. This still means that agent a can deduce, from $\mathcal{O}_a$ and A_{ab} that $C' \sqsupseteq D$ (resp. b can deduce that $C \sqsupseteq D'$ from $\mathcal{O}_b$ and A_{ab}). We call this *local entailment*.

Definition 2.8 (Local Correspondence Satisfiability)**.** Given two ontologies $\mathcal{O}_a$ and $\mathcal{O}_b$ and an interpretation $I = \langle \Delta, \cdot^I \rangle$ of $\mathcal{O}_a$. The interpretation I *locally satisfies* a correspondence between $C \in \mathcal{C}_a$ and $D \in \mathcal{C}_b$ (noted $\models_a$) if there exists an extension $\cdot^{I'}$ of $\cdot^I$ to $\mathcal{C}_a \cup \mathcal{C}_b$ such that:

$$I \models_a C \sqsubseteq D \text{ iff } C^{I'} \subseteq D^{I'}$$
$$I \models_a C \sqsupseteq D \text{ iff } C^{I'} \supseteq D^{I'}$$

As usual, a *local model* of an alignment for an ontology $\mathcal{O}$ is a model of $\mathcal{O}$ an extension of which locally satisfies all the correspondences of the alignment, and a *local model* of an ontology $\mathcal{O}$ for an set of alignments $\mathcal{E}$ is a model of $\mathcal{O}$ an extension of which locally satisfies all the correspondences in all alignments in $\mathcal{E}$. An alignment is *locally consistent* if it has a local model (otherwise locally inconsistent), an ontology is *locally consistent* for a set of alignments $\mathcal{E}$ if it has a local model for $\mathcal{E}$, and a correspondence γ is *locally entailed* for agent a by an alignment A if it is satisfied in all of its local models (noted $A \models_a \gamma$).

Therefore, given an interpretation I of an ontology $\mathcal{O}_a$, I locally satisfies a correspondence CRD between classes in $\mathcal{O}_a$ and another ontology $\mathcal{O}_b$ if we can extend I in such a way to interpret D so that CRD is satisfied. This means that, even though agent a does not have access to the ontology of agent

b, she can reason about the alignment between them. For example, it allows for agents to find out that there is no model compatible with an alignment. Consider for instance, $\{C(o), C \oplus C'\} \subseteq \mathcal{O}_a$ and $\{\langle C, D, \sqsubseteq\rangle, \langle C', D, \sqsupseteq\rangle\} \subseteq A_{ab}$, see Figure 2.2. There can be no extension to $\mathcal{C}_b$ of a model of $\mathcal{O}_a$ satisfying both correspondences. This is a good reason why agents may want to repair them.

In other words, local satisfiability depicts what the agents can reason about locally. The definition can be rewritten for covering how b interprets A_{ab}.

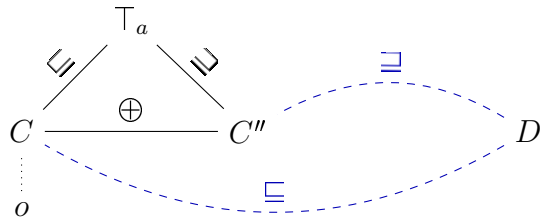

Figure 2.2: There is no local model for the ontology on the left and alignment consisting of the two correspondences to class D.

We split the alignments used in ARG into two distinct alignments using only the $\sqsupseteq$ relation. Any network of alignments may be rewritten with such conventions [47]. Moreover, the alignments have to follow specific constraints allowing to always find a single correspondence applicable for an object.

Definition 2.9 (ARG alignment)**.** An alignment $A_{ab} \subseteq \mathcal{C}_a \times \mathcal{C}_b \times \{\sqsupseteq\}$ between two ARG ontologies $\mathcal{O}_a$ and $\mathcal{O}_b$, is an ARG alignment if

- $\langle \top_a, \top_b, \sqsupseteq\rangle \in A_{ab}$,
- for each class $D \in \mathcal{C}_b$ there exists at most one class $C \in \mathcal{C}_a$ such that $\langle C, D, \sqsupseteq\rangle \in A_{ab}$.

Globally, we will consider as ARG states specific networks of aligned ontologies.

Definition 2.10 (ARG State)**.** An ARG state s, for a set $\mathcal{A}$ of agents, is the pair $s = \langle \{\mathcal{O}_a\}_{a\in\mathcal{A}}, \{A_{ab}\}_{a,b\in\mathcal{A},a\neq b}\rangle$, such that

- $\mathcal{O}_a$ is an ARG ontology associated to agent $a \in \mathcal{A}$;
- A_{ab} is an ARG alignment between $\mathcal{O}_a$ and $\mathcal{O}_b$.

We can define local and global consistency for ARG states, with respect to the local correspondence satisfiability relation.

Definition 2.11 (Local and Global Consistency ([47]))**.** An ARG state is said *locally consistent* for a set of agents $\mathcal{A}$ if it is locally consistent for each agent $a \in \mathcal{A}$, i.e. their ontology $\mathcal{O}_a$ is locally consistent for the set of alignments $\bigcup_{b \in \mathcal{A} \setminus \{a\}} A_{ab}$. An ARG state is said *globally consistent* for a set of agents $\mathcal{A}$ if there exists a tuple $\langle I_a \rangle_{a \in \mathcal{A}}$ of models for each ontology $\mathcal{O}_a$ which satisfies all the correspondences of all alignments.

2.1.1 Alignment Repair Game

There are many different ways in which such alignments have been used. It is possible to generate alignments beforehand and to use them, it is also possible to create them or compose them on the fly. ANEMONE [36] matches ontologies on the fly when necessary: in situations agents need to communicate, but cannot express themselves. In such situations, agents will exchange concept definitions or concept instances to reach a common understanding via alignments. Other approaches to compose alignments on the fly use argumentation to accept or reject correspondences from a library of alignments [69, 86].

There are two drawbacks to these approaches: it prevents agents to evolve their ontologies after the alignments have been composed, and such techniques consider the obtained alignment as fully correct and do not consider modifying or repairing it dynamically. In fact, the first situation can lead to the second: through evolving ontologies, alignments may become incorrect because an agent may adopt the meaning of parts of the ontology. Different techniques have been proposed to evolve alignments: gossiping amongst agents to reach global agreement [1], logical repair to enforce consistency [65, 73, 87], or prevention of logical violations to agents' ontologies via conservativity principles [66]. Some have been integrated with multi-agent systems via specific protocols [1, 80].

These repair techniques are developed independently of agent interaction. However, it may not be realistic nor desirable for agents to stop interacting until the repair is completed. To overcome this, *interaction-situated semantic alignment* was proposed [4]. This is an ontology matching algorithm embedded in the interaction protocols used by agents to communicate. Alignments are then induced whenever interactions are repeatedly successful and failing interactions lead to revision. This proposal was further advanced to repair alignments *through their use* and generalized to less constrained protocols [4, 31–33].

The Alignment Repair Game (ARG) [46] may be considered as belonging to this category of approaches. It takes inspiration from cultural evolution, which applies an idealized version of the theory of evolution to culture [75], where culture may be any shared artefact among a population such as food regime, or language. Work in cultural evolution is based on the observation of long-term behaviors of a population. With computers, cultural evolution can be modeled as dynamic systems and the quantitative findings can be compared to observations [29, 84]. Furthermore, they can also be used to explore small scale phenomena, such as the influence of the size of a population on artefact complexity [35]. Experiments in cultural evolution are performed through multi-agent simulation [5], where a population of agents interact and adapt their culture through a precisely defined protocol. In such experiments, agents repeatedly perform a random task, called a game, and their evolution is observed. The goal of this is to discover to what state agents converge and which properties hold or define that state. This has successfully been applied to the evolution of natural languages [91].

In ARG, experimental cultural evolution is applied to knowledge. More precisely, to the evolution of ontology alignments. It lets agents continuously play a communication game and systematically adapt their alignments when a failure occurs by offering a protocol designed for agents to evolve alignments between their ontologies through their use [46, 49]. The aim of ARG is to detect and repair mistakes in alignments whenever a communication failure occurs through application of the adaptation operators. The idea is that ultimately, by repeatedly playing ARG, the alignments will not cause failure any more.

Definition 2.12 (Alignment Repair Game)**.** The *Alignment Repair Game* is played a fixed number of rounds from an initial ARG state by a set of agents $\mathcal{A}$ with a common set $\mathcal{D}$ of object names from an ARG state s and for a chosen operator.

At each round of the game:

1. Two agents $a, b \in \mathcal{A}$ with $a \neq b$ and an object $o \in \mathcal{D}$ are picked at random.

2. Agent a asks agent b to which class in her ontology the object o belongs according to the alignment A_{ab}. Agent b answers the most specific class C_b that is identified via: $\mathcal{O}_b \models C_b(o)$, $\langle C_a, C_b, \sqsupseteq \rangle \in A_{ab}$, and $\nexists C'_b \neq C_b$ such that $\langle C'_a, C'_b, \sqsupseteq \rangle \in A_{ab}$ and $\mathcal{O}_b \models C'_b \sqsubseteq C_b$.

3. Agent a compares C_a with the object o. If $\mathcal{O}_a \models C_a(o)$, then the round is a success, else if $\mathcal{O}_a \not\models C_a(o)$ the round is a failure and an adaptation operator $\alpha[\langle C_a, C_b, \sqsupseteq \rangle, o]$ is applied to the alignment A_{ab}.

As an illustration of one ARG round consider Example 2.2 that will serve as a running example throughout this thesis.

Example 2.2 (Running example)**.** Let agent a and agent b play ARG where their ontologies $\mathcal{O}_a$ and $\mathcal{O}_b$ are described in Figure 2.1.

The initial alignment A_{ab} is represented by the blue dashed correspondences between classes of their ontologies. Now, consider two cases: the object ▲ and the object △. Let in both cases agent a ask agent b to which class the object belongs in her ontology so that it can be translated to $\mathcal{O}_a$ via the alignment. In both cases, agent b will answer $Small_b$ as both objects belong to this class in $\mathcal{O}_b$. However, while for the object ▲ the round would be successful (because through the alignment, $Small_b$ is translated to $Black_a$ and $\mathcal{O}_a \models Black_a(▲)$), for the object △ a failure is reached (because through the alignment, $Small_b$ is translated to $Black_a$, but $\mathcal{O}_a \models White_a(△) \wedge White_a \oplus Black_a$). In the latter case an adaptation operator $\alpha[\langle Black_a, Small_b, \sqsupseteq\rangle, △]$ has to be applied to the alignment A_{ab} (see Example 2.3).

The agent behavior, in this version of the game, is fully deterministic: given the ordered structure of the ontology and the uniqueness of the eligible correspondence in ARG alignments (granted by Definition 2.9 on page 18), the agent does not choose the correspondence to apply.

Given the failure of correspondence $\langle C_a, C_b, \sqsupseteq\rangle \in A_{ab}$ with object o, *adaptation operators* specify what the agents should do.

Definition 2.13 (Adaptation Operator)**.** An *adaptation operator* α is an alignment transformer $\alpha[c, o] : A_{ab} \mapsto A'_{ab}$ where A_{ab} and A'_{ab} are alignments, $c \in A_{ab}$ and o is an object.

We also write α for $\alpha[c, o]$ whenever c and o are clear from the context. For an ARG state $s = \langle\{\mathcal{O}_a\}_{a\in\mathcal{A}}, \{A_{ab}\}_{a,b\in\mathcal{A}}\rangle$, we also write $\alpha[c, o](s)$ for $\langle\{\mathcal{O}_a\}_{a\in\mathcal{A}}, \{A_{xy}\}_{(x,y)\in\mathcal{A}\times\mathcal{A}\setminus\{a,b\}} \cup \{\alpha[c, o](A_{ab})\}$. Again, whenever the correspondence and object are clear from the context, we also simply write $\alpha(s)$.

In [46, 49] the following adaptation operators $\alpha[\langle C_a, C_b, \sqsupseteq\rangle, o]$ are introduced:

- `delete`$[\langle C_a, C_b, \sqsupseteq\rangle, o]$: delete the correspondence $\langle C_a, C_b, \sqsupseteq\rangle$ from A_{ab};
- `add`$[\langle C_a, C_b, \sqsupseteq\rangle, o]$: in addition to `delete`$[\langle C_a, C_b, \sqsupseteq\rangle, o]$, add the correspondence $\langle msc_a(C_a), C_b, \sqsupseteq\rangle$ between C_b and the most specific superclass of C_a;

- `addjoin`$[\langle C_a, C_b, \sqsupseteq\rangle, o]$: in addition to `delete`$[\langle C_a, C_b, \sqsupseteq\rangle, o]$, add the correspondence $\langle mscc_a(o, C_a), C_b, \sqsupseteq\rangle$ between C_b and the most specific superclass of C_a that is compatible with the object o;

- `refine`$[\langle C_a, C_b, \sqsupseteq\rangle, o]$: in addition to `delete`$[\langle C_a, C_b, \sqsupseteq\rangle, o]$, add the correspondences $\langle C_a, C'_b, \sqsupseteq\rangle$ between C_a and all the most general subclasses C'_b of C_b that are not compatible with the object o (i.e. $\mathcal{O}_b \not\models C'_b(o)$) and which do not already have a correspondence $\langle C'_a, C'_b, \sqsupseteq\rangle \in A_{ab}$;

- `refadd`$[\langle C_a, C_b, \sqsupseteq\rangle, o]$: first apply `addjoin`$[\langle C_a, C_b, \sqsupseteq\rangle, o]$ and then apply `refine`$[\langle C_a, C_b, \sqsupseteq\rangle, o]$.

Formally, this amounts to:

$$
\begin{aligned}
\texttt{delete}[\langle C_a, C_b, \sqsupseteq\rangle, o](A_{ab}) &= A_{ab} \backslash \{\langle C_a, C_b, \sqsupseteq\rangle\} \\
\texttt{add}[\langle C_a, C_b, \sqsupseteq\rangle, o](A_{ab}) &= \texttt{delete}[\langle C_a, C_b, \sqsupseteq\rangle, o](A_{ab}) \\
&\quad \cup \{\langle msc_a(C_a), C_b, \sqsupseteq\rangle\} \\
\texttt{addjoin}[\langle C_a, C_b, \sqsupseteq\rangle, o](A_{ab}) &= \texttt{delete}[\langle C_a, C_b, \sqsupseteq\rangle, o](A_{ab}) \\
&\quad \cup \{\langle mscc_a(o, C_a), C_b \sqsupseteq\rangle\} \\
\texttt{refine}[\langle C_a, C_b, \sqsupseteq\rangle, o](A_{ab}) &= \texttt{delete}[\langle C_a, C_b, \sqsupseteq\rangle, o](A_{ab}) \\
&\quad \cup \ \{\langle C_a, C'_b, \sqsupseteq\rangle | C'_b \in mgcx_b(C_b, o) \text{ and} \\
&\qquad \nexists C'_a; \langle C'_a, C'_b, \sqsupseteq\rangle \in A_{ab}\} \\
\texttt{refadd}[\langle C_a, C_b, \sqsupseteq\rangle, o](A_{ab}) &= \texttt{addjoin}[\langle C_a, C_b, \sqsupseteq\rangle, o](A_{ab}) \\
&\quad \cup \texttt{refine}[\langle C_a, C_b, \sqsupseteq\rangle, o](A_{ab})
\end{aligned}
$$

As can be observed, some of the actions (`add`, `addjoin`) can only be performed by agent a, who is the only one to know $\mathcal{O}_a$, and some others (`refine`) can only be performed by agent b, for symmetric reasons. Hence, the implementation of these operators involves a gently asking b for performing `refine`, and part of `refadd`, upon failure.

The adaptation operators introduced in [46, 49] share two properties:

- *Safeness*: After applying the adaptation operator, if the same object is drawn, the same failure does not occur again (but maybe a different failure occurs);

- *Entailment* Every added correspondence by the adaptation operator was entailed by the failing (and removed) correspondence.

Where entailment is defined with respect to both ontologies and the alignment [45]. For example, for `add`, this holds because for any ontology $\mathcal{O}_a$ and alignment A_{ab} it is true that $C_a \sqsubseteq D_a \in \mathcal{O}_a$ and $\langle C_a, C_b, \sqsupseteq \in A_{ab}$ imply $D_a \sqsupseteq C_b$.

Furthermore, it is clear from the definition that every operator entails `delete`, and `refadd` entails `addjoin` and `refine`. The order of the actions that are performed by the adaptation operators does not matter. Figure 2.3 illustrates their effects.

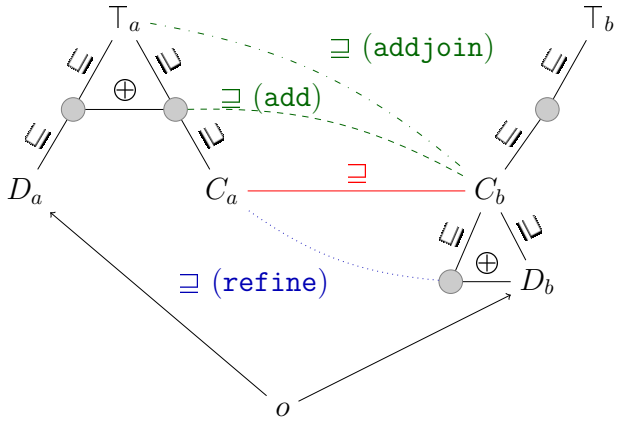

Figure 2.3: Schematic diagram of the deleted (red, solid) and added correspondences (green and dashed for `add`, green and dash-dotted for `addjoin`, blue and dotted for `refine`) by the different adaptation operators in ARG.

There is a link between the adaptation operators and the AGM framework of Belief Revision [3]. Comparing the adaptation operators to the AGM framework, safeness can be interpreted as the success postulate, and entailment the inclusion postulate. In fact, the adaptation operators discussed here are not revision operators but rather contraction operators. This follows from the entailment property and the fact that in Belief Revision Theory, only *closed sets* are considered. Closed sets X are sets such that whenever ϕ can be entailed from X, it holds that $\phi \in X$. Then, `addjoin` can be thought of as the minimal contraction operator amongst the ones discussed. This is because it only deletes the failing correspondence but not the correspondences entailed by it. However, on ARG, we do not consider closed sets but local correspondence satisfiability (Definition 2.8 on page 17). This differs in the sense that correspondences of the alignment are satisfied with respect to only one agent's ontology and the alignment, and not two agents' ontologies.

Consider an example of applying the adaptation operators.

Example 2.3 (Running example)**.** In the end of Example 2.2, an adaptation operator is applied to the alignment A_{ab}. This adaptation operator $\alpha[\langle Black_a, Small_b, \sqsupseteq\rangle, \triangle]$ deletes the initial correspondence and adds the following new correspondences to the alignment:

- `delete`: none
- `add`: $\langle \top_a, Small_b, \sqsupseteq\rangle$ $(msc_a(Black_a) = \top_a)$
- `addjoin`: $\langle \top_a, Small_b, \sqsupseteq\rangle$ $(mscc_a(\triangle, Black_a) = \top_a)$
- `refine`: $\langle Black_a, SS_b, \sqsupseteq\rangle$ $(mgcx_b(Small_b, \triangle) = SS_b)$
- `refadd`: $\langle \top_a, Small_b, \sqsupseteq\rangle$ and $\langle Black_a, SS_b, \sqsupseteq\rangle$

The alignment repair game modifies the situation from ARG state to ARG state, as expressed by Property 2.1.

Property 2.1 (Operators preserve ARG stateness [19])**.** Given an ARG state s and a failure of correspondence c with object o, then for each operator $\alpha \in \{\texttt{delete}, \texttt{add}, \texttt{addjoin}, \texttt{refine}, \texttt{refadd}\}$, $\alpha[c, o](s)$ is well-defined and is an ARG state.

2.1.2 Experimental Results

By playing ARG with different adaptation operators, they can be compared. This has been achieved experimentally in [46,49] for the adaptation operators discussed here. They have been evaluated with respect to four different measures:

- *Success rate* [91]: the ratio of success over games played;
- *Semantic precision and recall* [45]: the degree of correctness and completeness with respect to the reference alignment, the complete and correct alignment;
- *Incoherence rate* [74]: the proportion of incoherent correspondences in alignments taken one by one;
- *Convergence*: the (maximum) number of games needed to converge to complete success.

Of course, reference alignments are not known to the agents but can be generated and used for measuring the quality of the resulting alignments.

It was found that all the operators have a relatively high success rate, yet do not reach 100% precision, and convergence to successful communication [46, 49]. Of the different adaptation operators, `delete` converges more quickly than `add`, `addjoin`, `refine` and `refadd`. This can be explained because `delete` suppresses the correspondence, therefore removing the cause of the failure, while the other operators also add one or more correspondences, which may be incorrect.

However, quick convergence is not necessarily a guarantee for 'good' alignments. Even though agents may be able to communicate successfully, it does not mean that their alignments are close to the reference alignment. Indeed, `refadd`, followed by `addjoin` and `add`, show the highest semantic recall [49], meaning that their output alignments are more complete compared to the other adaptation operators. Furthermore, `add` shows particularly a low precision, high incoherence rate and slow convergence [49]. This is because `add` may add a correspondence to the alignment that will cause another failure when the same object is drawn in the future. This happens when the superclass of C_a may not be a class to which the object belong in agent a's ontology. In comparison, `addjoin` takes this into account and finds the lowest superclass of C_a that is compatible with the drawn object to add a correspondence to.

There are limitations to the adaptation operators: agents are restricted to the initial alignment (they do not introduce random new correspondences, only those entailed by the failing correspondence) and shadowed, false correspondence may remain undetected (they always use the most specific class in a correspondence) [49]. In [49], two "modalities" have been introduced to overcome these limitations called expansion and relaxation. With these modalities, it is possible to play ARG from empty alignments [48]. Both modalities improve the experimental measures discussed above but are out of the scope of this thesis so will not be further discussed here.

Despite the experimental results about ARG (agents converge towards successful communication through local corrective actions and improve their alignments [46, 49]), very little of the formal properties of the ARG agents or adaptation operators were assessed formally. This is the purpose of this thesis.

2.2 Dynamic Epistemic Logic

This thesis aims to study the theoretical properties of the Alignment Repair Game (ARG). For this, ARG is modeled in a logic based on Dynamic Epistemic Logic (DEL).

Dynamic Epistemic Logic is used to study knowledge, belief and other epistemic attitudes, studied in (multi-agent) epistemic logic [52,61,76], under model change, using formal languages and mathematical models [42]. More generally speaking, DEL is the study of modal logics of model change [23]. It has been widely used as a framework to model information flow in multi-agent systems and has been applied to communication [14, 42], belief revision [12] and agent interaction [13]. Therefore, it is a good candidate to formalize the Alignment Repair Game (ARG) and study its properties.

Often in DEL, logical puzzles are used to motivate and illustrate different model changing actions, for example the famous Muddy Children Puzzle [52], or the more recent puzzle regarding Cheryl's birthday [30].

DEL extends any given epistemic logic language (see [52] for a classic reference on epistemic logic) with one or more *modal operators* $\dagger\phi$, also called modalities, that describe model-transforming actions.

Definition 2.14 (Syntax of DEL)**.** Given a countable, non-empty set P of propositional letters and a finite, non-empty set $\mathcal{A}$ of agents, the *syntax*, $\mathcal{L}_{DEL}$, of (multi-agent) Dynamic Epistemic Logic is defined in the following way:

$$\phi ::= p \mid \phi \wedge \psi \mid \neg\phi \mid K_a\phi \mid B_a\phi \mid [\dagger\phi]\psi$$

where $p \in P$ is a proposition, K_a and B_a are the knowledge and belief operators for each agent a and $\dagger\phi$ with $\dagger \in \{!, \Uparrow, \uparrow\}$ the dynamic upgrades.

As usual, the connectives $\vee$ and $\rightarrow$, and the duals $\hat{K}_a, \hat{B}_a, \langle\dagger\phi\rangle$ can be defined by: $\phi \vee \psi$ iff $\neg(\neg\phi \wedge \neg\psi)$, $\phi \rightarrow \psi$ iff $\neg\phi \vee \psi$, $\hat{K}_a\phi$ iff $\neg K_a\neg\phi$, $\hat{B}_a\phi$ iff $\neg B_a\neg\phi$, and $\langle\dagger\phi\rangle$ iff $\neg[\dagger\phi]\neg\psi$.

We read the formula $K_a\phi$ as "agent a knows that ϕ is true", the formula $B_a\phi$ as "agent a believes that ϕ is true", the formula $[\dagger\phi]\psi$ as "if ϕ, then, after the dynamic upgrade $\dagger\phi$, ψ". Therefore,$\langle\dagger\phi\rangle\psi$ reads as "ϕ and, after the dynamic upgrade $\dagger\phi$, ψ". The dynamic upgrades are further discussed in Section 2.2.1 (page 31).

The standard semantics for $\mathcal{L}_{DEL}$ are given by means of Kripke models. A Kripke model consists of a Kripke frame equipped with a valuation function, where wR_av reads as "from world w, agent a considers v possible".

Definition 2.15 (Kripke Frame)**.** Given a finite, non-empty set $\mathcal{A}$ of agents, a *Kripke frame* for $\mathcal{A}$ is a pair $\mathfrak{F} = \langle W, (R_a)_{a\in\mathcal{A}}\rangle$ where

- W is a non-empty set of *worlds*;
- and $(R_a)_{a\in\mathcal{A}}$ is a set of binary relations over W indexed by the agents: $R_a \subseteq W \times W$ for each $a \in \mathcal{A}$.

Definition 2.16 (Kripke Model)**.** Given a countable, non-empty set P of propositional letters and a finite, non-empty set $\mathcal{A}$ of agents, a *Kripke model* for $\mathcal{A}$ and P is a pair $\mathcal{M} = \langle \mathfrak{F}, V \rangle$ where

- $\mathfrak{F}$ is a DEL frame;
- and $V : P \to \mathcal{P}(W)$ is a propositional valuation function mapping propositions to sets of worlds in which that proposition is true.

A *pointed Kripke model* is a pair $\langle \mathcal{M}, w \rangle$ where $\mathcal{M}$ is a Kripke model and $w \in W$.

By convention, $W^{\mathcal{M}}$, $R_a^{\mathcal{M}}$ and $V^{\mathcal{M}}$ are used to refer to the components of $\mathcal{M}$, but we omit the superscript $\mathcal{M}$ if it is clear from the context which model we are concerned with. We also write $w \in \mathcal{M}$ write to mean $w \in W^{\mathcal{M}}$. Furthermore, we write $V_w(p) = 1$ to denote that $w \in V(p)$, and $V_w(p) = 0$ to denote that $w \notin V(p)$.

When we draw models, we write, inside the worlds, p to denote that the valuation of p at that world is 1 and $\bar{p}$ to denote that the valuation is 0. Relations wR_av are represented by drawing an arrow from w to v ($w \to v$) with the label a. We also double circle a world w to denote that the pointed model is $\mathcal{M}, w$. Consider the following example of a Kripke model.

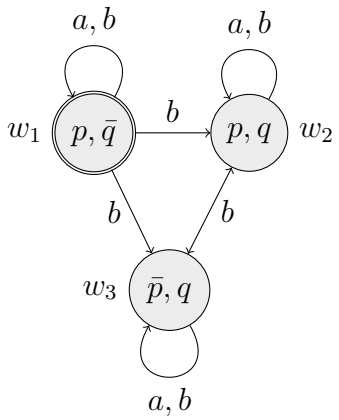

Figure 2.4: The Kripke model $\mathcal{M}$ for two agents a and b as defined in Example 2.4.

Example 2.4 (Running Example)**.** Let $\mathcal{M} = \langle W, (R_a)_{a \in \{a,b\}}, V \rangle$ be the Kripke model depicted in Figure 2.4 where $W = \{w_1, w_2, w_3\}$, the relations are $R_a = \{\langle w, w \rangle\}_{w \in W}$, $R_b = \{\langle w_1, w_2 \rangle, \langle w_1, w_3 \rangle, \langle w_2, w_3 \rangle, \langle w_3, w_2 \rangle\} \cup$

$\{\langle w, w\rangle\}_{w\in W}$ and the valuation is $V_{w_1}(p) = V_{w_2}(p) = V_{w_2}(q) = V_{w_3}(q) = 1$, $V_{w_1}(q) = V_{w_3}(p) = 0$.

Thus, for each agent a, we have a relation R_a specifying which worlds the agent considers and, possibly, in what order. The relation is then used to define knowledge and beliefs of agents: agent a knows something if and only if it is true at all the worlds agent a considers (in any direction) and agent a believes something if and only if it is true at the maximal worlds (forward) with respect to the relation for a. Knowledge and belief can also be defined with respect to the epistemic ($\sim_a$) and doxastic relations ($\rightarrow_a$).

Definition 2.17 (Epistemic and Doxastic Relation)**.** Let $\langle W, \{R_a\}_{a\in\mathcal{A}}, V\rangle$ be a DEL model for a set $\mathcal{A}$ of agents, then the epistemic relation $\sim_a$ is defined as:

$$w \sim_a v \;\text{ iff }\; w\ (R_a \cup R_a^{-1})^*\ v \tag{2.2}$$

And the doxastic relation $\rightarrow_a$ is defined as:

$$w \rightarrow_a v \;\text{ iff }\; v \in Max_{R_a}|w|_a \tag{2.3}$$

where R^* is the transitive closure of any relation R and $|w|_a$ is the *information cell* (or *accessible cell*) of agent a at state w and is defined by:

$$|w|_a = \{v \in W \mid w \sim_a v\} \tag{2.4}$$

Then an agent a knows ϕ if it holds at all the worlds reached via $\sim_a$ and she believes ϕ if it holds at all the worlds reached via $\rightarrow_a$. Consider again the model in Example 2.4. We can now draw the epistemic and doxastic relations, see Figure 2.5.

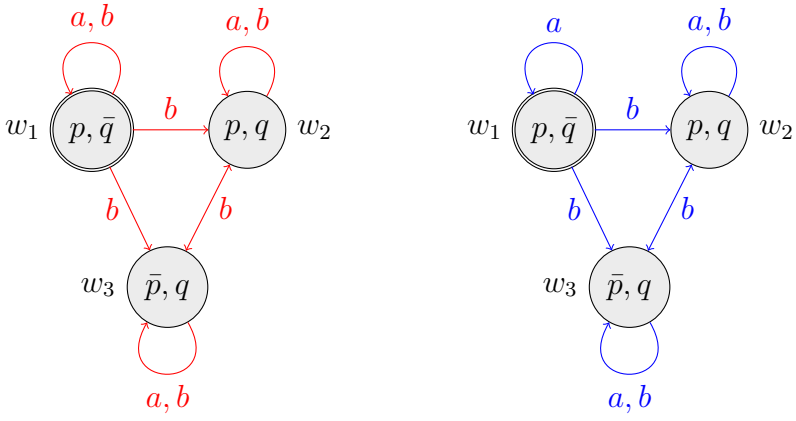

Figure 2.5: The epistemic (in red on the left) and doxastic (in blue on the right) relations for the DEL model $\mathcal{M}$ defined in Example 2.4.

In Figure 2.5, everything that is true in all the worlds accessible through the epistemic relations is the *knowledge* of an agent, and everything that is true in all the worlds accessible through the doxastic relations is the *belief* of an agent

Satisfiability is defined with respect to pointed models $\langle \mathcal{M}, w \rangle$.

Definition 2.18 (Satisfiability for DEL)**.** *Satisfiability* for Dynamic Epistemic Logic by a pointed model $\langle \mathcal{M}, w \rangle$ is defined in the following way:

$$
\begin{array}{ll}
\mathcal{M}, w \vDash p & \text{iff } w \in V(p) \\
\mathcal{M}, w \vDash \phi \wedge \psi & \text{iff } \mathcal{M}, w \vDash \phi \text{ and } \mathcal{M}, w \vDash \psi \\
\mathcal{M}, w \vDash \neg\phi & \text{iff } \mathcal{M}, w \nvDash \phi \\
\mathcal{M}, w \vDash K_a\phi & \text{iff } \forall v \text{ s.t. } w \sim_a v : \mathcal{M}, v \vDash \phi \\
\mathcal{M}, w \vDash B_a\phi & \text{iff } \forall v \text{ s.t. } w \rightarrow_a v : \mathcal{M}, v \vDash \phi \\
\mathcal{M}, w \vDash [\dagger\phi]\psi & \text{iff } \mathcal{M}^{\dagger\phi}, w \vDash \psi
\end{array}
$$

where $\dagger \in \{!, \Uparrow, \uparrow\}$ are model transformers $\dagger\phi : \mathcal{M} \mapsto \mathcal{M}^{\dagger\phi}$ whose domain and range is the set of Kripke models (defined in Section 2.2.1 on page 32).

We also write $\dagger_1\phi; \dagger_2\psi$ for the sequence of upgrades $\dagger_1\phi$ and then $\dagger_2\psi$ meaning that first $\dagger_1\phi$ is applied and then $\dagger_2\phi$, i.e. the model $\mathcal{M}^{\dagger_1\phi;\dagger_2\psi}$ is defined as $(\mathcal{M}^{\dagger_1\phi})^{\dagger_2\psi}$.

We may use $\bot$ to denote $p \wedge \neg p$ for any proposition $p \in P$, which is false on all (non-empty) models, and likewise $\top$ to denote $\neg\bot$, which is true on all (non-empty) models.

Example 2.5 (Running Example)**.** In the model in Figure 2.4, we have, among others, that $\mathcal{M}, w_1 \vDash K_a(p \wedge \neg q)$, $\mathcal{M}, w_1 \nvDash K_b p$, $\mathcal{M}, w_1 \nvDash K_b q$ and $\mathcal{M}, w_1 \vDash B_b q$.

We say that a set of formulas is consistent if there is a pointed model satisfying all formulas of the set. Otherwise, a set of formulas is inconsistent. In the following, a formula ϕ is said to be a consequence of a set of formulas Γ (written $\Gamma \vDash \phi$) if every pointed model $\langle \mathcal{M}, w \rangle$ satisfying all formulas of Γ, also satisfies ϕ.

If we consider all Kripke models, the set of valid formulas obtained via the satisfiability relation constitutes a logic. One example of such a valid formula is the axiom K (or actually, axiom schemata): $K_a(\phi \rightarrow \psi) \rightarrow (K_a\phi \rightarrow K_a\psi)$. Furthmore, restricting the class of Kripke frames to specific kinds of relations corresponds to adding specific axioms to the logic. For example, the class of transitive frames is characterized by the axiom $K_a\phi \rightarrow K_aK_a\phi$, see table 2.1.

Name	Axiom	Frames	Logical Property
K	$K_a(\phi \to \psi) \to (K_a\phi \to K_a\psi)$		distributivity
D	$K_a\phi \to \neg K_a\neg\phi$	serial	consistency
T	$K_a\phi \to \phi$	reflexive	truth
4	$K_a\phi \to K_aK_a\phi$	transtive	positive introspection
5	$\neg K_a\phi \to K_a\neg K_a\phi$	Euclidean	negative introspection

Table 2.1: The axiom schemata for DEL and the corresponding properties on frames and in the logic.

Most often, when dealing with knowledge, relations are assumed to be equivalence relations which corresponds to the logic $S5$ ($K + T + 4 + 5$ in Table 2.1). Similarly, when dealing with belief, the logic $KD45$ is standard. This provides quite strong notions of knowledge and belief. For example, concerning knowledge, whatever is known also has to be true, if an agent knows something she knows that she knows it and if an agent does not know something, she knows that she does not know it. In particular the latter two, positive and negative introspection, are controversial in philosophy [83].

In this thesis, we will be dealing both with knowledge and belief, requiring a variation of Kripke models based on a *plausibility relation* $\leqslant_a$ [10]. These are called *plausibility models*. The relation $w \leqslant_a v$ reads as "agent a considers v more plausible than w". In the following, these will be the standard and also referred to as models, or DEL models.

We first define relational properties that will be used when introducing plausibility frames.

Definition 2.19 (Relation Properties)**.** Given a non-empty set of worlds W and an accessibility relation $R_a \subseteq W \times W$, we say that R_a is

reflexive	iff	$\forall w \in W$:	wR_aw
transitive	iff	$\forall w, v, u \in W$:	$wR_av, vR_au \Rightarrow wR_au$
locally connected	iff	$\forall w, v \in W$:	$v \in \lvert w\rvert_a \Rightarrow wR_av$ or vR_aw
well-founded	iff	$\forall S \subseteq W, S \neq \varnothing$:	$Max_a(S) \neq \varnothing$

where $Max_a(S) = \{w \in S \mid \forall v \in S : wR_av\}$.

A relation R_a is called a *preorder* if R_a is both reflexive and transitive.

Therefore, a relation R_a is locally connected if anything in the accessible cell of an agent is connected by R_a, and it is well-founded if every subset of W has a set of maximal elements with respect to R_a.

Definition 2.20 ((Plausibility/DEL) Frame)**.** Given a finite, non-empty set $\mathcal{A}$ of agents, a *plausibility frame* for $\mathcal{A}$ is a pair $\mathfrak{F} = \langle W, (\leqslant_a)_{a\in\mathcal{A}}\rangle$ where

- W is a non-empty set of *worlds*, and
- $(\leqslant_a)_{a\in\mathcal{A}}$ is a set of plausibility relations $\leqslant_a \subseteq W \times W$, one for each agent, that are well-founded, locally connected preorders;

Definition 2.21 ((Plausibility/DEL) Model)**.** Given a countable, non-empty set P of propositional letters and a finite, non-empty set $\mathcal{A}$ of agents, a *plausibility model* for $\mathcal{A}$ and P is a pair $\mathcal{M} = \langle \mathfrak{F}, V\rangle$ where

- $\mathfrak{F}$ is a plausibility frame;
- and $V : P \to \mathcal{P}(W)$ is a propositional valuation function mapping propositions to sets of worlds in which that proposition is true.

A *pointed plausibility model* is a pair $\langle \mathcal{M}, w\rangle$ where $\mathcal{M}$ is a plausibility model and $w \in W$.

We can define $\sim_a$ and $\to_a$ with respect to $\leqslant_a$ in the same way it was conducted for R (Equations 2.2 and 2.3 on page 28). It then follows from the properties of $\leqslant_a$ that the epistemic relations $\sim_a$ are reflexive, transitive and symmetric, and the doxastic relations $\to_a$ are transitive, serial and Euclidean. Therefore they satisfy the usual properties of knowledge and belief, $S5$ ($K + T + 4 + 5$ in Table 2.1 on page 30) and $KD45$ ($K + D + 4 + 5$ in Table 2.1 on page 30), respectively [42]. The resulting logic is also called *(multi-agent) epistemic-doxastic logic*.

2.2.1 Announcements, Radical and Conservative Upgrades

The modal operators, defining model transforming actions, make DEL different from epistemic logic. These operators allow formulas to be interpreted across models. For example, if A is such a modal operator, formulas of the form $[A]\phi$ express that after applying action A to a model $\mathcal{M}$ to obtain $\mathcal{M}^A$, ϕ is true in $\mathcal{M}^A$. This means that DEL shifts from a static semantics of truth taking place in an individual model to a dynamic semantics of truth taking place across models.

The first approach to add such modal operators to logic is [82], first published in [81] in 1989. This logic is nowadays called Public Announcement Logic (PAL) and extends epistemic logic with a modal operator to describe *public announcements* $!\phi$. Public announcements, or in short announcements,

are the model transforming actions that delete all the worlds from the model where the announced statement is false. I.e. $!\phi$ deletes all $\neg\phi$-worlds. As a consequence, in the obtained model, ϕ will be true everywhere and therefore also common knowledge to the agents.

DEL generalizes the ideas by [82] to other modal operators because it is not always feasible that what is communicated is *public*, *truthful* and *trusted*, as is true for public announcements. Most notably are *radical upgrades* $\Uparrow\phi$, *conservative upgrades* $\uparrow\phi$ [12] and *private announcements* $!_G\phi$ [7] (discussed in Section 2.2.2). For the first two, DEL needs to be considered extended with a belief operator: these upgrades change the beliefs of agents, instead of knowledge.

Radical and conservative upgrades have been discussed under various names in the context of Belief Revision (for example the AGM framework [54]), e.g. in [85, 95]. They have been formalized for DEL in an attempt to bridge DEL with Belief Revision in [12]. We can think of radical and conservative upgrades as communication by a trusted, but fallible source. For this reason, as model transformers they do not delete any worlds from the model, but instead, change the plausibility relations of agents so that the carried information is pushed to the top of the relation. The difference between the two is that in case of radical upgrades, the information carried is highly trusted, whereas for conservative upgrades, it is 'just' trusted. Fixed points of announcements, radical and conservative upgrades have been investigated in [9].

Definition 2.22 (Model Transformer)**.** A *model transformer* $\dagger\phi$ is a function $\dagger\phi : \mathcal{M} \mapsto \mathcal{M}^{\dagger\phi}$, applying a certain action to $\mathcal{M}$ to obtain $\mathcal{M}^{\dagger\phi} = \langle W^{\dagger\phi}, (\leqslant_a^{\dagger\phi})_{a\in\mathcal{A}}, V^{\dagger\phi}\rangle$. We consider three model transformers $!\phi$, $\Uparrow\phi$ and $\uparrow\phi$ that are defined as follows, with $||\phi||_{\mathcal{M}}$ denoting the set of worlds in which ϕ is true, i.e. $||\phi||_{\mathcal{M}} = \{w \in W \mid \mathcal{M}, w \vDash \phi\}$:

Announcement ($!\phi$) deletes all '$\neg\phi$'-worlds from the model, i.e. $W^{!\phi} = ||\phi||_{\mathcal{M}}$, $w \leqslant_a^{!\phi} v$ iff $w \leqslant_a v$ and $w, v \in W^{!\phi}$, $V^{!\phi}(p) = V(p) \cap ||\phi||_{\mathcal{M}}$;

Radical upgrade ($\Uparrow\phi$) makes all ϕ worlds more plausible than all $\neg\phi$ worlds, and within these two zones, the old ordering remains. I.e. $W^{\Uparrow\phi} = W$, $w \leqslant_a^{\Uparrow\phi} v$ iff $v \in ||\phi||_{\mathcal{M}}$ and $w \in ||\neg\phi||_{\mathcal{M}}$ or else if $w \leqslant_a v$, and $V^{\Uparrow\phi}(p) = V(p)$;

Conservative upgrade ($\uparrow\phi$) makes the best 'ϕ'-worlds more plausible than all other worlds, while the old ordering on the rest of the worlds remains. I.e. $W^{\uparrow\phi} = W$, $w \leqslant_a^{\uparrow\phi} v$ iff either $v \in Max_{\leqslant_a}(|w|_a \cap ||\phi||_{\mathcal{M}})$ or $w \leqslant_a v$, $V^{\Uparrow\phi}(p) = V(p)$.

Let us look more in detail at the modal operators. Announcements remove worlds that do not make the carried information true, radical upgrades push all worlds that make the carried information true on top of the plausibility order and conservative upgrades do the same but only with the 'best worlds' in which the carried information is true. The modal operators preserve the properties of DEL models [12]. However, there are restrictions to them. For example, announcements $!\phi$ can only be applied to a pointed model $\langle \mathcal{M}, w \rangle$ if ϕ is true at w. Similarly, $\Uparrow p$ and $\uparrow p$ do not alter the model if $\neg p$ is true.

To illustrate the modal operators, consider again the example before, where $!p$, $\Uparrow p$ and $\uparrow p$ are applied.

Example 2.6 (Running Example)**.** Let $\mathcal{M}$ be the DEL model defined in Example 2.4. Then $\mathcal{M}^{!p}$, $\mathcal{M}^{\Uparrow p}$ and $\mathcal{M}^{\uparrow p}$ are depicted in Figure 2.6. It holds that, among others, $\mathcal{M}^{!p}, w \vDash K_b p$, $\mathcal{M}^{\Uparrow p}, w \vDash B_b p$ and $\mathcal{M}^{\uparrow p}, w \vDash B_b p$.

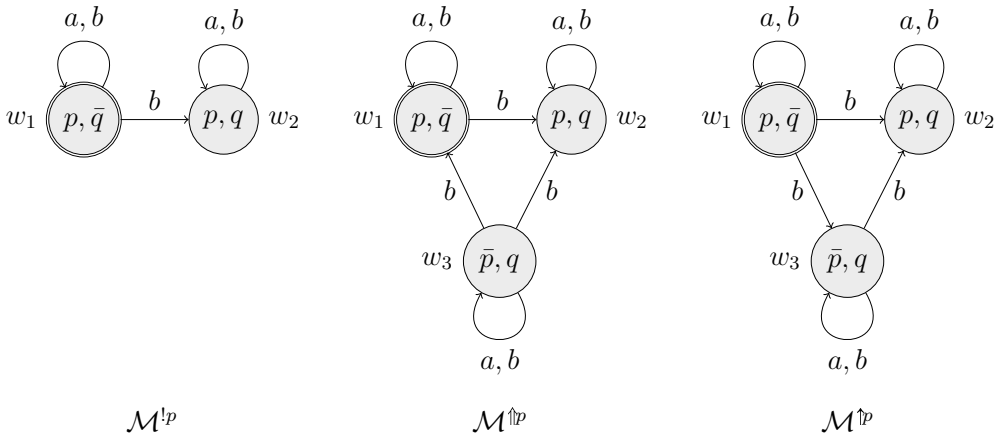

Figure 2.6: The announcement $!p$ (left), radical upgrade $\Uparrow p$ (middle) and conservative upgrade $\uparrow p$ (right) applied to the DEL model $\mathcal{M}$ as defined in Example 2.4.

When applying model transformers, the knowledge and beliefs of agents are changing. In Example 2.6, after announcing p ($!p$) the agents come to know p (and they know that they both know) and after a radical ($\Uparrow p$) or conservative upgrade with p ($\uparrow p$) agent b comes to believe p. The difference between $\Uparrow p$ and $\uparrow p$ become apparent when considering a strong version of belief: an agent a *strongly believes* ϕ if ϕ is considered possible by a and furthermore all ϕ-worlds accessible by a are more plausible than all $\neg\phi$-worlds accessible by a, see [9] for a formal definition.

Interestingly, the modal operators discussed so far do not actually add something new to the language. Indeed, the expressivity is the same as for an epistemic-doxastic logic: any sentence involving an announcement, radical or conservative upgrade can be reduced to one without [12, 81]. We present the *reduction axioms* for announcements, the ones for radical and conservative upgrades can be found in [12].

Definition 2.23 (Reduction Axioms for $!\phi$)**.** The following formulas are the *reduction axioms* for announcements $!\phi$.

- $[!\phi]p \leftrightarrow (\phi \rightarrow p)$
- $[!\phi]\neg\psi \leftrightarrow (\phi \rightarrow \neg[!\phi]\psi)$
- $[!\phi](\psi_1 \wedge \psi_2) \leftrightarrow ([!\phi]\psi_1 \wedge [!\phi]\psi_2)$
- $[!\phi]K_a\psi \leftrightarrow (\phi \rightarrow K_a(\phi \rightarrow [!\phi]\psi))$
- $[!\phi][!\chi]\psi \leftrightarrow [!\phi \wedge [!\phi]\chi]\psi$

The reduction axioms are valid on any model [55, 81] Therefore, a complete axiomatization for DEL is given by the axioms of a logic for knowledge and belief of choice (typically S5 for the knowledge operator and KD45 for the belief operator [22]), the reduction axioms [12, 55, 81], $K_a\phi \rightarrow B_a\phi$ (knowledge implies belief) and $B_a\phi \rightarrow K_aB_a\phi$ (positive introspection of belief). Still it must be noted though that there are different ways to axiomatize DEL, as has been shown for Public Announcement Logic (PAL) [96].

When working with DEL models, a useful and important notion is that of *bisimulation*. It formalizes when two models are semantically equivalent.

Definition 2.24 (Bisimulation)**.** Let two DEL models $\mathcal{M} = \langle W, (\leqslant_a)_{a\in\mathcal{A}}, V\rangle$ and $\mathcal{M}' = \langle W', (\leqslant'_a)_{a\in\mathcal{A}}, V'\rangle$ be given for a finite, non-empty set of agents $\mathcal{A}$.

A relation $Z \subseteq W \times W'$ is a *bisimulation* if and only if for all $(w, w') \in Z$ the following three conditions hold:

- [**Propositional agreement**] $V(w) = V'(w')$;
- [**Forth**] For every agent $a \in \mathcal{A}$ and for every $v \in W$ such that $w \leqslant_a v$ there exists a $v' \in W'$ such that $w' \leqslant'_a v'$ and $(v, v') \in Z$;
- [**Back**] For every agent $a \in \mathcal{A}$ and for every $v' \in W'$ such that $w' \leqslant'_a v'$ there exists a $v \in W$ such that $w \leqslant_a v$ and $(v, v') \in Z$.

Two pointed models $\langle \mathcal{M}, w\rangle$ and $\langle \mathcal{M}', w'\rangle$ are *bisimilar* if and only if there is a bisimulation Z such that $(w, w') \in Z$.

In DEL, the semantic notion of bisimulation coincides with the statement that models satisfy the same formulas. That is, semantic and syntactic equivalence are, indeed, equivalent [42].

Theorem 2.1 ([42])**.** If two pointed DEL models for the same set of agents are bisimilar, then they satisfy the same formulas.

2.2.2 Dynamic Epistemic Logic with Event Models

In the previous section, three ways have been discussed and formalized to change the knowledge and beliefs of agents. In this section, we look at a generalization of these upgrades called *event models*, or *action models*, introduced in [7]. Event models are relational structures that allow us to talk about the dynamics of information in the same way that Kripke models formalize static information. The general idea is to think of event models as Kripke models, but instead of consisting of worlds, we consider it to be consisting of a set of *events*, and instead of a valuation a *precondition* is defined. Like DEL models, the relational structure specifies which events the agents can tell apart.

Event models can be used to describe a variety of informational events: from public announcements to more subtle communication containing privacy, misleading or suspicion. For example, information may be shared in secret, hidden completely (other agents do not observe the communication) or partially (other agents observe the communication, but not what is communicated) from others. A classic example of this is the coin-toss example.

Example 2.7 (Coin Toss)**.** Agents a, b, c play a coin-toss. Assume agent c throws a coin, catches it in her palm and fully covers it before anybody (including agent c) can see on which side the coin has landed, so that nobody sees the upper face of the coin. The event model for such situation is depicted in Figure 2.7.

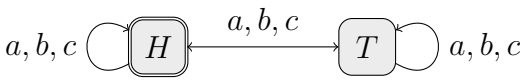

Figure 2.7: The event model for the coin toss as described in Example 2.7.

Example 2.8 (Coin Toss)**.** Let the situation be as in Example 2.7, but now, let us assume that agent c was cheating: she took a look at the coin (it

was heads up) before covering it and nobody noticed this. Assume also that agent c knows that a and b do not suspect anything. The event model for this situation is depicted in Figure 2.8.

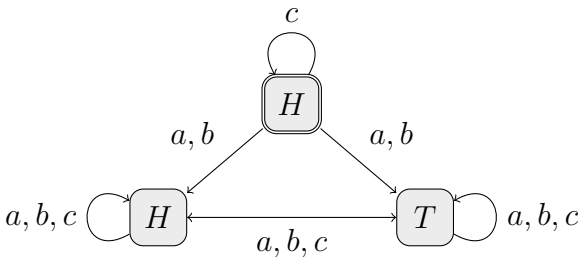

Figure 2.8: The event model for the coin toss in which agent c cheats as described in Example 2.8.

Examples 2.7 and 2.8 show the complexity added by event models. These are not the only situations we can imagine: what happens if a sees that c is cheating, without c noticing that a caught her? Or if c tells b she cheated?

To formalize these situations, we first introduce an alternative language to the one discussed in Section 2.2: the language of Epistemic Action Logic (EAL) [7].

Definition 2.25 (Syntax of EAL)**.** Given a countable, non-empty set P of propositional letters and a finite, non-empty set $\mathcal{A}$ of agents, the *syntax*, $\mathcal{L}_{EAL}$, of (multi-agent) Epistemic Action Logic (EAL) is defined in the following way.

$$\phi ::= p \mid \phi \wedge \psi \mid \neg\phi \mid K_a\phi \mid B_a\phi \mid [\langle \mathcal{E}, e \rangle]\phi$$

where $p \in P$ is a proposition, K_a and B_a are the knowledge and belief operators for each agent a, and $\langle \mathcal{E}, e \rangle$ are pointed event models.

Frames and models of EAL are equivalent to DEL frames and models (Definitions 2.20 and 2.21 on pages 31, 31), with a plausibility relation for each agent that is a well-founded, locally connected preorder.

Event models for EAL provide a relational structure to dynamic upgrades [7]. With this we can capture public announcements, but also private announcements to a group $G \subseteq \mathcal{A}$, meaning that only the agents in G "observe" the announcement and the other agents are unaffected. Formally, an event model is like a Kripke model but instead of worlds we consider *events* and instead of a valuation a *precondition* is defined.

Definition 2.26 (Event Model)**.** Let P be a countable, non-empty set of propositional letters and let $\mathcal{A}$ be a finite, non-empty set of agents. An *event model* for EAL is a triple $\mathcal{E} = \langle E, (R_a)_{a \in \mathcal{A}}, pre \rangle$ where

- E is a non-empty, finite set of *events*;
- $(R_a)_{a \in \mathcal{A}} \subseteq E \times E$ are the *accessibility relations* on E, one for each agent $a \in \mathcal{A}$;
- $pre : E \to \mathcal{L}_{EAL}$ is a *precondition function* assigning to each event a formula ϕ.

A *pointed event model* is a pair $\langle \mathcal{E}, e \rangle$ where $\mathcal{E}$ is an event model and $e \in E$.

We will also write pre_e for $pre(e)$.

When drawing a pointed event model $\langle \mathcal{E}, e \rangle$, events are drawn as squares to distinguish them from worlds and e is double-squared to emphasize the point of reference.

To determine what happens if an event model $\mathcal{E}$ takes place on a DEL model $\mathcal{M}$, their *product update* $\mathcal{M} \otimes \mathcal{E}$ is computed [7].

Definition 2.27 (Product Update)**.** Let $\mathcal{M} = \langle W, (\leqslant_a)_{a \in \mathcal{A}}, V \rangle$ be a EAL model and $\mathcal{E} = \langle E, (R_a)_{a \in \mathcal{A}}, pre \rangle$ be an event model. Their *product update*, denoted by $\mathcal{M} \otimes \mathcal{E}$, is the triple $\langle W^{\mathcal{M} \otimes \mathcal{E}}, (\leqslant_a^{\mathcal{M} \otimes \mathcal{E}})_{a \in \mathcal{A}}, V^{\mathcal{M} \otimes \mathcal{E}} \rangle$ defined by:

- $W^{\mathcal{M} \otimes \mathcal{E}} = \{\langle w, e \rangle \in W \times E \mid \mathcal{M}, w \vDash pre(e)\}$
- $\langle w, e \rangle \leqslant_a^{\mathcal{M} \otimes \mathcal{E}} \langle w', e' \rangle$ iff $\langle w, e \rangle, \langle w', e' \rangle \in W^{\mathcal{M} \otimes \mathcal{E}}$, $w \leqslant_a w'$ and $e R_a e'$
- $V^{\mathcal{M} \otimes \mathcal{E}}(p) = \{\langle w, e \rangle \in W \times E \mid w \in V(p)\}$

The product update $\mathcal{M} \otimes \mathcal{E}$ is the result of the events $e \in E$ happening at the worlds $w \in W$ whenever w satisfies the precondition $pre(e)$. The precondition therefore serves as a selection to which worlds an event may be applied. For example, if $pre(e) = \phi$, then the event e may only be applied to worlds w that make ϕ true. Then, if e is the sole event of $\mathcal{E}$, this means that the worlds falsifying ϕ are deleted from the product update. In the following we also refer to the events e such that $\mathcal{M}, w \vDash pre(e)$ as the events that *can be applied* to w.

Because agents may observe the event differently, the accessibility relations R_a express how the different agents observe the event. This determines which relations remain in the product update from the initial epistemic model. Hence, for a to have access to a world in the product update, there

needs to be a $\leqslant_a$-arrow between the corresponding worlds and a R_a-arrow between their corresponding events.

Satisfiability for EAL extends that of DEL (Definition 2.18 on page 29) with a clause for pointed event models replacing the clause for $[\dagger\phi]\psi$. Satisfiability for pointed event models is determined with respect to the product update.

Definition 2.28 (Satisfiability for Events)**.** *Satisfiability* for EAL extends satisfiability for DEL (Definition 2.18 on page 29), replacing the last clause by:

$$\mathcal{M}, w \vDash [\langle \mathcal{E}, e \rangle]\psi \text{ iff } \mathcal{M}, w \vDash pre(e) \textit{ implies that } \mathcal{M} \otimes \mathcal{E}, \langle w, e \rangle \vDash \psi$$

Public announcement

The event model for a public announcement $!\phi$ consists of a single event with precondition ϕ and a reflexive relation for all agents.

Definition 2.29 (Public Announcement)**.** The pointed event model for the public announcement $!\phi$ is $\langle \mathcal{E}, e_\phi \rangle$ where the event model is defined as $\mathcal{E}_{!\phi} = \langle \{e_\phi\}, (I_a)_{a \in \mathcal{A}}, pre \rangle$ and $pre(e_\phi) = \phi$, see Figure 2.9.

Figure 2.9: The event model $\mathcal{E}_{!\phi}$ for a public announcement $!\phi$.

Indeed, the event model for public announcements is equivalent to the model transformer $!\phi$ in definition 2.18: only the worlds satisfying the precondition, namely ϕ, remain in the resulting model, while the accessibility relations to and from these worlds are equivalent to the relations in the initial model. In other words, $\neg\phi$-worlds are deleted.

Consider again the running example, applying a public announcement.

Example 2.9 (Running Example)**.** Let $\mathcal{M}$ be the DEL model defined in Example 2.4. Then $\mathcal{M} \otimes \mathcal{E}_{!p}$ is the DEL model resulting from applying $\mathcal{E}_{!p}$ to $\mathcal{M}$, see Figure 2.10.

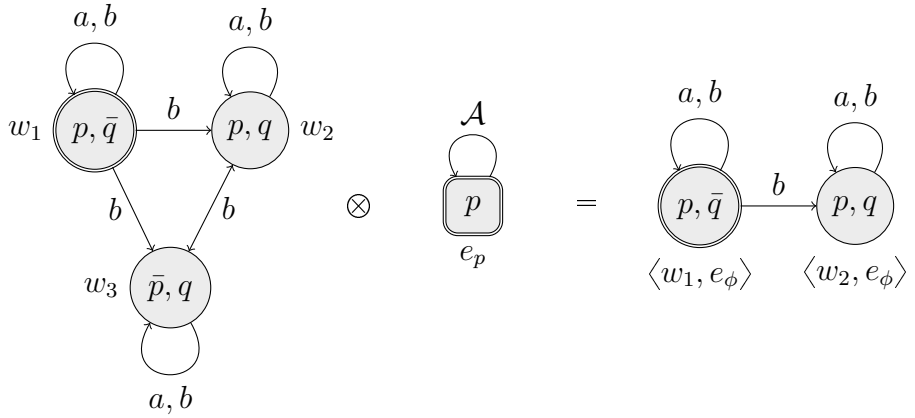

Figure 2.10: The event model $\mathcal{E}_{!p}$ for a public announcement of p applied to the DEL model $\mathcal{M}$ as defined in Example 2.4.

Private announcement

Private announcements are announcements that areonly received by a subset G of agents. Therefore, the event model for a private announcement consists of two events, one where the precondition is the announcement and relations are reflexive for the agents of G (this event defines the reference set), and another where the precondition is true, that is reached from the first event for all agents not in G [7].

Definition 2.30 (Private Announcement)**.** The pointed event model for the fully private announcement $!_G\phi$ to a group $G \subseteq \mathcal{A}$ is $\langle \mathcal{E}_{!_G\phi}, e_\phi \rangle$ where the event model is defined as $\mathcal{E} = \langle \{e_\phi, e_\top\}, (R_a)_{a \in \mathcal{A}}, pre \rangle$ such that for $a \in G$, $R_a = \{\langle e_\phi, e_\phi \rangle, \langle e_\top, e_\top \rangle\}$ and otherwise $R_a = \{e_\phi, e_\top\} \times \{e_\top\}$, $pre(e_\phi) = \phi$ and $pre(e_\top) = \top$, see Figure 2.11.

After public announcements $!\phi$, all agents come to know that ϕ, whereas after private announcements $!_G\phi$, all agents in G come to know that ϕ, i.e. $[!_G\phi]K_a\phi$ holds for all $a \in G$ [7].

Consider again the running example, applying a private announcement $!_{\{b\}}p$ only to agent b.

Example 2.10 (Running Example)**.** Let $\mathcal{M}$ be the EAL model defined in Example 2.4. Then $\mathcal{M} \otimes \mathcal{E}_{!_{\{b\}}p}$ is the EAL model resulting from applying $\mathcal{E}_{!_{\{b\}}p}$ to $\mathcal{M}$, see Figure 2.12.

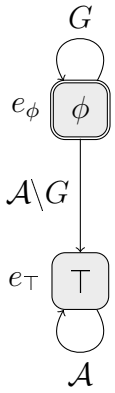

Figure 2.11: The event model $\mathcal{E}_{!_G\phi}$ for a private announcement $!_G\phi$ to a group of agents G.

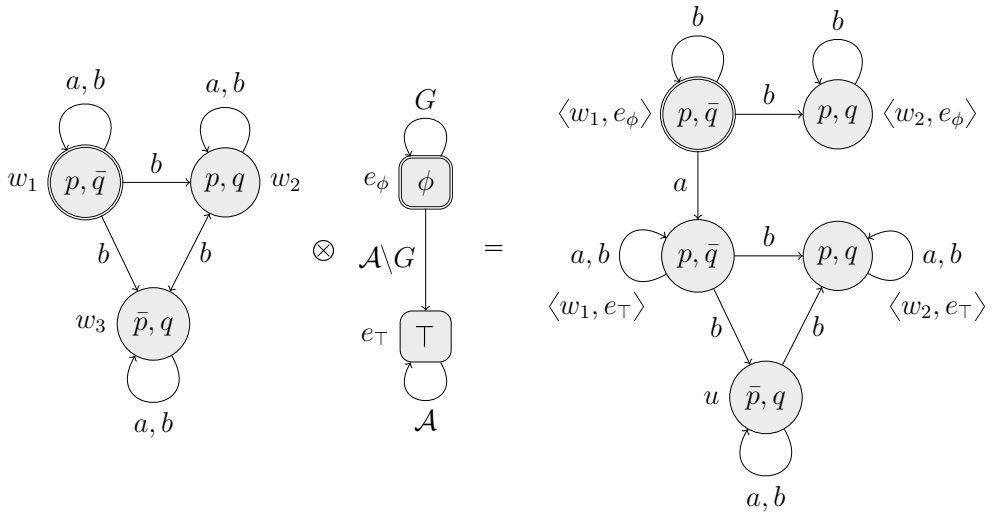

Figure 2.12: The event model $\mathcal{E}_{!_{\{b\}}p}$ for a private announcement of p to agent b applied to the EAL model $\mathcal{M}$ as defined in Example 2.4.

Whereas public announcements preserve the structures of EAL models that were stipulated in Definition 2.21 (page 31)– the plausibility relations are well-founded, locally connected preorders – this is not true for private announcements. To ensure that the event takes place privately, the agents $b \notin G$ do not have access to e_ϕ in Definition 2.30 (page 39). I.e. there is no reflexive relation for these agents at e_ϕ. This means that, in the product

update, the worlds 'created' with e_ϕ do not have reflexive relations for $b \notin G$. As a consequence, the resulting models are no longer EAL models. To prevent this, typically event models are required to be such that they preserve the properties of the model structures [7]. Then, it holds that applying an event to a EAL model yields a EAL model. In the situation here, this requirement boils down to requiring the relations R_a in event models to satisfy the same properties as the plausibility relations $\leqslant_a$ in EAL models [7].

Of course, this restriction greatly limits the event models that can be applied. For example, it excludes private announcements, as is clear from Example 2.10. Another way to deal with this would be to weaken the logic. We will come back to this in Section 5.10 (page 111) when we introduce awareness.

Event models with postconditions

Event models as described here have been generalized in [14] to accommodate factual change by introducing *postconditions*.

Definition 2.31 (Event Model with Postconditions)**.** Let P be a countable, non-empty set P of propositional letters and let $\mathcal{A}$ be a finite, non-empty set of agents. An *event model with postconditions* for EAL is a quadruple $\mathcal{E} = \langle E, (R_a)_{a\in\mathcal{A}}, pre, post\rangle$ where

- E is a non-empty, finite set of *events*,
- $(R_a)_{a\in\mathcal{A}} \subseteq E \times E$ are the *accessibility relations* on E, one for each agent $a \in \mathcal{A}$,
- $pre : E \to \mathcal{L}_{DEL}$ is a *precondition function* assigning to each event a formula ϕ, and
- $post : E \times P \to \mathcal{L}_{DEL}$ is a *postcondition function* assigning to each event e and proposition p a formula $post(e, p) \in \mathcal{L}_{DEL}$.

A *pointed event model with postconditions* is a pair $\langle \mathcal{E}, e\rangle$ where $\mathcal{E}$ is an event model with postconditions and $e \in E$.

We will also write pre_e for $pre(e)$ and $post_e(p)$ for $post(e, p)$.

Postconditions are used to alter the valuation function of a model. They assign to each event $e \in E$ and proposition $p \in P$ a formula $post(e, p)$ so that for each $\langle w, e\rangle \in W^{\mathcal{M}\otimes\mathcal{E}}$ in the product update $\langle w, e\rangle \in V^{\mathcal{M}\otimes\mathcal{E}}(p)$ if and only if $\mathcal{M}, w \vDash post(e, p)$. Hence, it makes the truth of p in the product model equivalent to the truth of $post(e, p)$ in the original model.

Definition 2.32 (Product Update for Events with Postcondtions)**.** Let $\mathcal{M} = \langle W, (\leqslant_a)_{a\in\mathcal{A}}, V\rangle$ be a EAL model and $\mathcal{E} = \langle E, (R_a)_{a\in\mathcal{A}}, pre, post\rangle$ be an event model with postconditions. Their *product update*, denoted by $\mathcal{M}\otimes\mathcal{E}$, is the triple $\langle W^{\mathcal{M}\otimes\mathcal{E}}, (\leqslant_a^{\mathcal{M}\otimes\mathcal{E}})_{a\in\mathcal{A}}, V^{\mathcal{M}\otimes\mathcal{E}}\rangle$ defined by:

- $W^{\mathcal{M}\otimes\mathcal{E}} = \{\langle w, e\rangle \in W \times E \mid \mathcal{M}, w \vDash pre(e)\}$
- $\langle w, e\rangle \leqslant_a^{\mathcal{M}\otimes\mathcal{E}} \langle w', e'\rangle$ iff $\langle w, e\rangle, \langle w', e'\rangle \in W^{\mathcal{M}\otimes\mathcal{E}}$, $w \leqslant_a w'$ and eR_ae'
- $V^{\mathcal{M}\otimes\mathcal{E}}(p) = \{\langle w, e\rangle \in W \times E \mid \mathcal{M}, w \vDash post(e, p)\}$

In [14] postconditions are restricted to be finite: $post(e, p) = p$ for all but finitely many pairs $\langle e, p\rangle \in E \times P$. This is to ensure event models to be finite objects.

2.3 Logic and Multi-Agent Systems

Dynamic Epistemic Logic (DEL) is a good candidate to formalize the Alignment Repair Game (ARG) and study its properties. Alternatively, one may start with a language for translating and relating different logic-based framework in a distributed manner such as with the Distributed Ontology, Modeling and Specification Language (DOL) [79]. DOL is supported by reasoning engines for heterogeneous reasoning and therefore provides a way for capturing agents ontologies and alignments. However, unlike DEL-like approaches, DOL has no means of capturing knowledge and beliefs of agents, nor their communication.

There are also other approaches to use logic to model multi-agent systems, or games [63]. For example, in [34], a framework for minimizing disagreements among beliefs is introduced where beliefs are associated with points on an undirected graph and revision takes place with respect to beliefs of "neighbors". The same approach is also used for Belief Revision Games [88], but here graphs may be directed, operators can be applied iteratively and agents can also drop beliefs. They then study what properties (standard AGM style revision postulates [3]) are needed to achieve consistency preservation, agreement preservation, convergence, stability, etc. Both of these experimental approaches are somewhat equivalent to the idea in [71], where a logic for the dynamics of belief change in a *community* is introduced and agents update their beliefs whenever a certain threshold is reached. This threshold is calculated with respect to the beliefs of friends: an agent is strongly influenced to believe p if and only if all her friends believe p and she is weakly influenced to believe p if and only if no friends believe $\neg p$. This

logic can then be used to study how the community can reach a stable belief state.

However, the connection between alignment repair and logic has not yet been made, even though alignment repair can be viewed as a belief revision strategy. Beyond the specific case of ARG, this thesis shows a methodology to model alignment repair in Dynamic Epistemic Logic by encoding ontologies and alignments as knowledge and beliefs, and capture the adaptation operators as dynamic upgrades for changing beliefs of agents. Then, properties such as correctness, completeness and redundancy can be defined and studied.

2.4 Awareness

Possible world semantics that are widely used for (Dynamic) Epistemic Logic assume that agents can talk about any formula of the language. In other worlds, that agents are *aware* of all these formulas. This is a strong idealization and leaves out a distinction between *implicit* and *explicit* knowledge, or changes in awareness. With regards to the model of ARG, it causes agents to use a single, common vocabulary to represent their knowledge and beliefs, violating the principle of preserving heterogeneity on ARG. Here, we provide a brief summary of the work on the notion of awareness relevant to this thesis.

Awareness has been first formalized in logic by Fagin and Halpern [51] as one way to solve the problem of logical omniscience [61]. They wonder how agents can say that they know or believe something about a proposition p if p is a concept they are completely unaware of. As a consequence, the interpretation of $K_a\phi$ is changed from "agent a knows ϕ" to "agent a *implicitly* knows ϕ", what relates to the knowledge an agent *could* eventually get. Another notion of knowledge called *explicit knowledge* is defined as a combination of implicit knowledge and awareness [51]. The logic they introduce, called Awareness Logic, extends the language of Epistemic Logic with an operator $A_a\phi$ that reads as "agent a is aware of ϕ" that is interpreted with respect to an awareness function assigning to each world and each agent a set of formulas. This function acts as a filter: extracting explicit knowledge from implicit knowledge.

2.4.1 Raising Awareness

With this notion of awareness, an interest has arisen to study the dynamics of awareness, e.g. [15, 37–39, 57, 59, 60, 77]. Particularly interesting to this

thesis, in [37, 38] the work by Fagin and Halpern [51] is extended to account for dynamic awareness, where the dynamic part is modeled by a bisimulation quantification on structures. Event models for changing awareness following this approach are defined in [39, 40]. Modalities to change awareness have been introduced in [15, 38], called the *consider* operation and *drop* operation, that extend or reduce the scope of the awareness function. A complete dynamic epistemic logic of awareness is then defined and the operations are generalized to multi-agent situations where awareness changes may occur privately [15, 40].

These works on awareness have mainly concentrated on awareness of the truth value of a statement not on awareness of the statement itself. This is because they are still based on total valuation functions and only awareness is considered as a partial function. This means that raising awareness comes equiped with disclosing an 'underlying' truth value of the proposition awareness is raised of [37, 38]. Even though it is possible in these logics to define a raising awareness awareness operation that does not lead to agents acquires knowledge or believe of the proposition or its negation awareness is raised of [40], i.e. to make the agents solely ignorant, it is required that this is determined in advance. That is, the valuations of the proposition awareness is raised of are already given, but are 'invisible' to the agents. Therefore, two problems remain: it disables agents to openly evolve their signatures when encountering new information and future evolutions of agents' knowledge and beliefs, and now also awareness, are bound by the initial setting.

In the notion of awareness introduced in this paper, awareness is implicitly used to define what an agent knows and believes, for which no awareness operator $A\phi$ or awareness function is required. Instead, awareness arises from the use of *partial valuation functions* and *weakly reflexive relations*. This means that agents may use different signatures, but we also tackle the other two problems: their signatures can openly evolve via raising awareness operations causing knowledge, belief and awareness to evolve without any prior defined way of how this evolution might take place. In addition, awareness is completely disconnected from truth: raising awareness does not imply disclosing its truth values.

Novel in this thesis is the connection between partial valuation functions and awareness. Partial valuation functions have been introduced for (Dynamic) Epistemic Logic in [62, 64, 94], where worlds of the models are equipped with partial, instead of total, valuation functions that interpret the propositions as true, false or *undefined*. This offers a more natural way to deal with growth of information by extending the models with it rather than reducing or re-organizing models, what is the typical approach in standard Dynamic Epistemic Logic to model changes in information [42]. In the latter

case, certainty grows in parallel with new information, but partial valuation functions allow for agents to grow their awareness in parallel with new information. Yet, a link to awareness, and therefore the possibility to agents to use different, dynamic signatures to represent their knowledge and beliefs, has not yet been established.

We extend the work by [58,62,64,94] in three ways: (1) we replace reflexivity by *weak reflexivity*, (2) we consider a different clause for the falsification of conjunctions and (3) we make a connection between partial valuations and awareness of agents. Weak reflexivity enables us to model how agents with different signatures, other agents are unaware of, interact. The new clause for falsification of conjunctions enforces that both conjuncts must belong to the domain of the valuation functions at the world considered in order to be false. This ensures that agents can only know that a conjunction "p and q" is false, if they aware of both p and q and know that at least one of the two conjuncts is false. Finally, the connection with awareness, gives a novel semantics for awareness and unawareness of agents in which becoming aware of a proposition and learning its truth value are two independent acts and in which agents can extend their signatures when learning new information from the environment or from other agents.

2.4.2 Forgetting awareness

Reverse modalities and operations have been studied throughout the history of logic: for example, the AGM model for belief revision considers expansion as well as contraction [3] and temporal logics are defined in function of *future* and *past* modalities [24, 25]. Forgetting was first studied in propositional and first order logics from a perspective of knowledge representation in [70]. Reasoning about knowledge under variable forgetting has been studied [92], and forgetting has been linked to uniform interpolation [44, 68, 97] and bisimulation invariance [72].

In epistemic logic, the notion of forgetting was studied in a number of ways. In [11] a 'forgetting knowledge' update is considered with the effect $\neg K\phi \wedge \neg K\neg\phi$: after knowledge forgetting ϕ, the agent would neither know ϕ nor $\neg\phi$. Considering awareness, forgetting has been defined in [37, 41]. In particular, [41] proposed a dynamic epistemic logic with an epistemic operator K and a dynamic modal operator $[Fg(p)]$ so that formula $[Fg(p)]\phi$ means that after the agent forgets his knowledge about p, ϕ is true.

Besides the theoretical motivation, there is also a practical motivation to consider forgetting coming from multi-agent systems. In such systems, agents use different ontologies and alignments to represent their knowledge and beliefs. During communication, they may encounter a counter-example

to their alignments that they revise accordingly [89, 93]. However, since they do not need specific examples to communicate successfully, they do not store them.

2.5 Conclusion

The Alignment Repair Game (ARG) has been introduced as a experimental framework of cultural knowledge evolution [46, 49]. It defines a specific protocol for agents, with different knowledge representations, to evolve their alignments: the agents play a communication game and whenever a failure is reached, they apply an adaptation operator. Through simulations with large numbers of games, properties such as convergence and success rate can be studied and established [46, 49].

However, they are not sufficient to understand the logical properties of cultural knowledge evolution. Therefore, we will introduce a logic based on Dynamic Epistemic Logic to model ARG and assess its properties.

Chapter 3

A Logical Model for the Alignment Repair Game

The Alignment Repair Game (ARG) models and assesses cultural knowledge evolution experimentally. It can be used to study how agents evolve their knowledge in situated environments and properties such as convergence can be established through experiments and simulations. However, these experiments and simulations are not sufficient to understand the logical properties of cultural knowledge evolution, whether the adaptation operators applied in ARG are formally correct, complete or redundant.

This chapter bridges experimental cultural knowledge evolution and a theoretical model of cultural knowledge evolution in logic. In particular, the formal properties of ARG are investigated: how the adaptation operators applied in ARG compare to the mechanisms in logic for agents to evolve their knowledge and beliefs. This is achieved by introducing an extension of Dynamic Epistemic Logic called Dynamic Epistemic Ontology Logic in which ontologies and alignments can be embedded. Then, a translation from ARG to DEOL is defined that (a) encodes ontologies, (b) translates agents' ontologies and alignments as knowledge and beliefs and (c) translates adaptation operators as announcements and conservative upgrades. Correctness, completeness and redundancy are defined with respect to this translation and it is proven that all but one adaptation operator are formally correct, all adaptation operators are incomplete and some adaptation operators are partially redundant. Finally, two interpretations of this failure to satisfy formal properties are discussed.

3.1 Dynamic Epistemic Ontology Logic

We introduce Dynamic Epistemic Ontology Logic (DEOL) as an extension of Dynamic Epistemic Logic where the propositions are object classifications ($C(x)$) and class relations ($C \sqsubseteq D$ and $C \oplus D$) of a Description Logic language. This is a minimal Description Logic language, suitable for representing ARG states. As seen in Chapter 2, ontologies and alignments can be completely encoded using these two features (Definitions 2.2 and 2.9 on pages 12, 18).

Definition 3.1 (Syntax of DEOL)**.** Given a countable, non-empty set $\mathcal{C}$ of class names, a countable, non-empty set $\mathcal{D}$ of object names, and a finite, non-empty set $\mathcal{A}$ of agents, the *syntax*, $\mathcal{L}_{DEOL}$, of (multi-agent) DEOL is defined in the following way:

$$\phi ::= C(o) \mid CRD \mid \phi \wedge \psi \mid \neg\phi \mid K_a\phi \mid B_a\phi \mid [\dagger\phi]\psi$$

$$R \in \{\sqsubseteq, \sqsupseteq, \oplus\},\ \dagger \in \{!, \Uparrow, \uparrow\}$$

where $C, D, \top \in \mathcal{C}$, $o \in \mathcal{D}$, K_a and B_a are the knowledge and belief operators for agent a and $\dagger\phi$ with $\dagger \in \{!, \Uparrow, \uparrow\}$ are the dynamic upgrades.

The connectives $\rightarrow$ and $\vee$ and the duals $\hat{K}_a, \hat{B}_a, \langle\dagger\phi\rangle$ for K_a, B_a and $[\dagger\phi]$, respectively, are defined as usual, analogous to the case of DEL.

Frames of DEOL are the same as (plausibility) frames of DEL. Of course, they can also be defined for more general for relations R_a not satisfying the properties of $\leqslant_a$, but we are interested in knowledge and beliefs of agents and therefore the use of $\leqslant_a$ is standard.

Definition 3.2 (DEOL Frames)**.** Given a finite, non-empty set $\mathcal{A}$ of agents, *frame* of (multi-agent) DEOL is a pair $\mathfrak{F} = \langle W, (\leqslant_a)_{a\in\mathcal{A}}\rangle$ where

- W is a non-empty set of worlds, and
- $(\leqslant_a)_{a\in\mathcal{A}} \subseteq W \times W$ are the plausibility relations on W, one for each agent, that are well-founded, locally connected preorders.

The difference between DEL and DEOL arises when we turn to models. In DEOL, the valuation function that assigns true or false to propositions in DEL is replaced by a pair $\langle\Delta, I\rangle$ where Δ is the domain, representing a set of objects, and I is an interpretation function that assigns to each world a function that interprets classes as set of objects of the domain.

Definition 3.3 (DEOL Model)**.** Given a countable, non-empty set $\mathcal{C}$ of class names, a countable, non-empty set $\mathcal{D}$ of object names, and a finite, non-empty set $\mathcal{A}$ of agents, a *model* of (multi-agent) DEOL is a triple $\mathcal{M} = \langle \mathfrak{F}, \Delta, I \rangle$ where

- $\mathfrak{F}$ is a DEOL frame,
- Δ is the domain of interpretation, and
- I is an *interpretation function* such that $I(w) = \cdot^{I_w}$ and $\cdot^{I_w}$ assigns to object names $o \in \mathcal{D}$ an element of the domain Δ ($\cdot^{I_w} : \mathcal{D} \to \Delta$), and to class names $C \in \mathcal{C}$ a subset of Δ ($\cdot^{I_w} : \mathcal{C} \to \mathcal{P}(\Delta)$), where it holds that $\top^{I_w} = \Delta$.

A *pointed DEOL model* is a pair $\langle \mathcal{M}, w \rangle$ where $\mathcal{M}$ is a DEOlm model and $w \in \mathcal{M}$.

Satisfiability for DEOL extends that of DEL (Definition 2.18 on page 29) by replacing propositions p by instance classifications $C(o)$ and class relations $C \sqsubseteq D$ and $C \oplus D$. Again, satisfiability is considered with respect to pointed models.

Definition 3.4 (Satisfiability for DEOL)**.** *Satisfiability* for Dynamic Epistemic Ontology Logic by a pointed model $\langle \mathcal{M}, w \rangle$ is defined in the following way:

$$
\begin{array}{ll}
\mathcal{M}, w \vDash C(o) & \text{iff } o^{I_w} \in C^{I_w} \\
\mathcal{M}, w \vDash C \sqsubseteq D & \text{iff } C^{I_w} \subseteq D^{I_w} \\
\mathcal{M}, w \vDash C \equiv D & \text{iff } C^{I_w} = D^{I_w} \\
\mathcal{M}, w \vDash C \oplus D & \text{iff } C^{I_w} \cap D^{I_w} = \varnothing \\
\mathcal{M}, w \vDash \phi \wedge \psi & \text{iff } \mathcal{M}, w \vDash \phi \text{ and } \mathcal{M}, w \vDash \psi \\
\mathcal{M}, w \vDash \neg\phi & \text{iff } \mathcal{M}, w \nvDash \phi \\
\mathcal{M}, w \vDash K_a\phi & \text{iff } \forall v \text{ s.t. } w \sim_a v : \mathcal{M}, v \vDash \phi \\
\mathcal{M}, w \vDash B_a\phi & \text{iff } \forall v \text{ s.t. } w \to_a v : \mathcal{M}, v \vDash \phi \\
\mathcal{M}, w \vDash [\dagger\phi]\psi & \text{iff } \mathcal{M}^{\dagger\phi}, w \vDash \psi
\end{array}
$$

where $C, D \in \mathcal{C}$, $o \in \mathcal{D}$, and $\dagger \in \{!, \Uparrow, \uparrow\}$ are model transformers: announcements ($!$), radical ($\Uparrow$) and conservative upgrades ($\uparrow$).

The model transformers for DEOL are defined as for DEL (Definition 2.22 on page 32). Notice that, again, we use $\mathcal{M}, w \vDash C \between D$ and say that "C and D overlap" whenever $\mathcal{M}, w \nvDash C \oplus D$, i.e. whenever $C^{I_w} \cap D^{I_w} \neq \varnothing$.

As for DEL, a formula ϕ is a consequence of a set of formulas Γ (written $\Gamma \models \phi$) if every pointed model $\langle \mathcal{M}, w \rangle$ satisfying all formulas of Γ, also satisfy ϕ.

By switching to propositions in the form $C(o)$ and CRD, we arrive at new axiom schemata that are valid on all models of DEOL, for example:

$$C \sqsubseteq C \tag{3.1}$$

$$\neg C \oplus C \tag{3.2}$$

$$\big(C(x) \wedge C \sqsubseteq D\big) \rightarrow D(x) \tag{3.3}$$

$$\big(C(x) \wedge C \oplus D\big) \rightarrow \neg D(x) \tag{3.4}$$

$$\big(C(x) \wedge D(x)\big) \rightarrow \big(C \sqsubseteq D \vee D \sqsubseteq C\big) \tag{3.5}$$

$$\big(C \sqsubseteq C' \wedge C' \sqsubseteq D\big) \rightarrow C \sqsubseteq D \tag{3.6}$$

$$\big(C \sqsubseteq C' \wedge C' \oplus D\big) \rightarrow C \oplus D \tag{3.7}$$

3.2 Translation

To investigate the formal properties of the adaptation operators applied by agents in ARG whenever a communication failure occurs, a translation is introduced that:

- maps ARG states, the set of ontologies and alignments used by agents in the system, to axioms of DEOL (τ), and
- maps adaptation operators to announcements and conservative upgrades (δ).

Figure 3.1 illustrates the translation applied to an ARG state s and an adaptation operator α. This figure will be used throughout this chapter as a tool to visualize the formal properties: correctness, completeness and redundancy.

3.2.1 ARG States as DEOL Axioms (τ)

An ARG state is a pair consisting of the ontologies and alignments used by agents in the system it describes. Ontologies are formal theories that serve as knowledge representations of the agents. It specifies, given a set of objects $\mathcal{D}$ and a set of classes $\mathcal{C}$, what relations hold between them according to a certain agent. Therefore, an ontology explains what the agent *knows*.

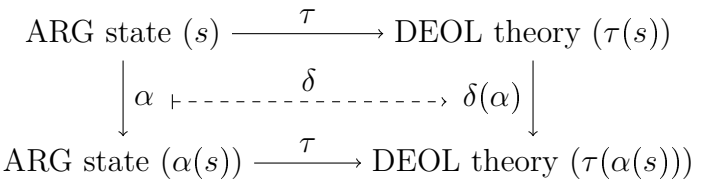

Figure 3.1: The translation from ARG states (s) to DEOL theories (τ) and from adaptation operators (α) to dynamic upgrades (δ).

Alignments define correspondences between entities of two agents' ontologies. These alignments may or may not be correct, leading to agents to evolve them. This means that alignments are not truthful and should therefore be treated accordingly as the *beliefs* of agents.

The translation τ from ARG states to axioms of DEOL is therefore defined as follows: any statement ϕ in the ontology $\mathcal{O}_a$ of agent a translates to knowledge of that agent, and any correspondence γ in the alignment A_{ab} between agent a's and b's ontologies translates to beliefs of both these agents.

Definition 3.5 (Translation τ)**.** The translation τ from ARG states to DEOL theories is defined by:

$$\begin{aligned}\tau(\langle\{\mathcal{O}_a\}_{a\in\mathcal{A}}, \{A_{ab}\}_{a,b\in\mathcal{A}}\rangle) = & \bigcup_{a\in\mathcal{A}}\{K_a\phi \mid \phi \in \mathcal{O}_a\} \\ & \cup \bigcup_{a,b\in\mathcal{A}}\{B_a\gamma \wedge B_b\gamma \mid \gamma \in A_{ab}\}\end{aligned}$$

Example

To illustrate the translation with an example, let us consider again the ARG state $\langle\{\mathcal{O}_a, \mathcal{O}_b\}, \{A_{ab}\}\rangle$ depicted in Figure 2.1, where the ontology of agent a is on the left, the ontology of agent b on the right and the alignment between their ontologies is drawn with blue, dashed lines.

When applying the translation τ to $\langle\{\mathcal{O}_a, \mathcal{O}_b\}, \{A_{ab}\}\rangle$, the following DEOL axioms are acquired:

$K_a(SB_a \sqsubseteq Black_a)$
$K_a(LB_a \sqsubseteq Black_a)$
$K_a(SW_a \sqsubseteq White_a)$
$K_a(LW_a \sqsubseteq White_a)$
$K_a(Black_a \sqsubseteq T_a)$
$K_a(White_a \sqsubseteq T_a)$

$K_a(SB_a(\blacksquare))$
$K_a(SB_a(\blacktriangle))$
$K_a(LB_a(\blacksquare))$
$K_a(LB_a(\blacktriangle))$
$K_a(SW_a(\square))$
$K_a(SW_a(\triangle))$

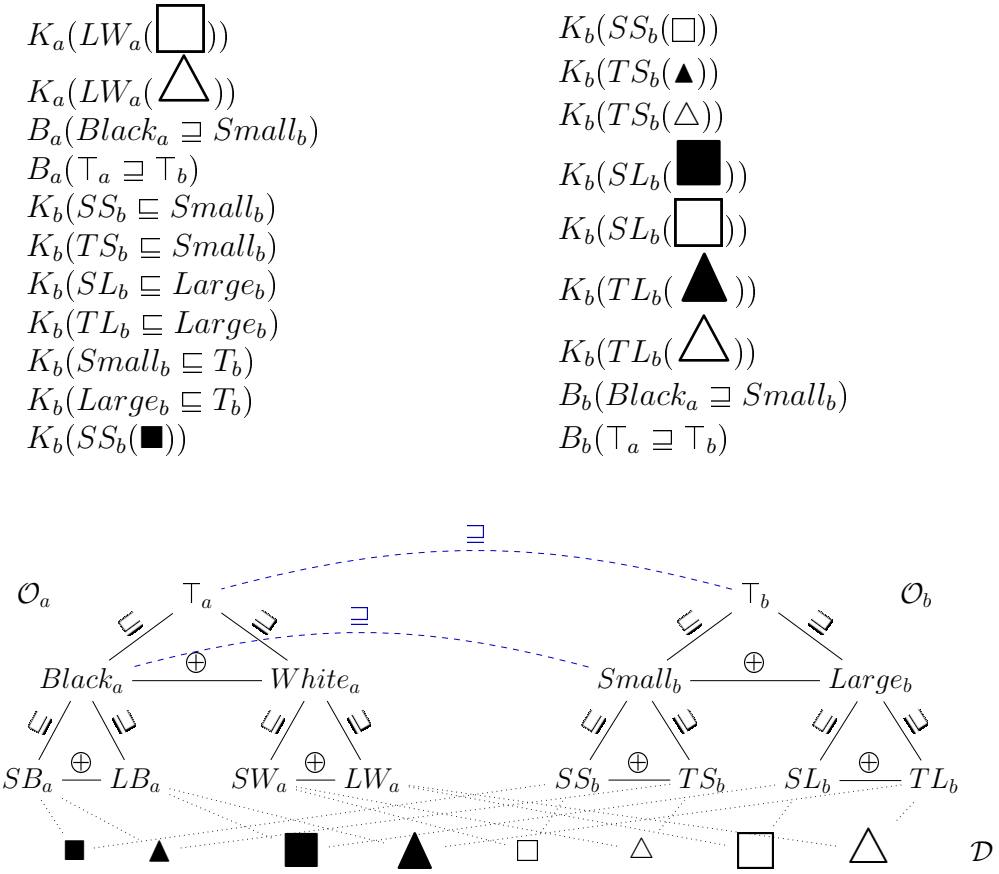

Figure 2.1: The ARG state $s = \langle\{\mathcal{O}_a, \mathcal{O}_b\}, \{A_{ab}\}\rangle$. (repeated from page 13)

The knowledge and beliefs of the agents can also be depicted graphically, like Figure 3.2 shows for agent a. In this figure, anything known by agent a is drawn in black and anything believed by agent a is drawn in blue.

Kripke models of $\tau(\langle\{\mathcal{O}_a, \mathcal{O}_b\}, \{A_{ab}\}\rangle)$ are such that everything that is known by the agents is true at each world, and whatever the agents believe is true in the most plausible worlds. This means that for any $K_a\phi$ or $K_b\phi$, ϕ must hold at any world of any model of $\tau(\langle\{\mathcal{O}_a, \mathcal{O}_b\}, \{A_{ab}\}\rangle)$, and for any $B_a\psi$ or $B_b\psi$, ψ is true at the most plausible worlds for agents a and b of any model of $\tau(\langle\{\mathcal{O}_a, \mathcal{O}_b\}, \{A_{ab}\}\rangle)$.

Before defining the translation of adaptation operators and discuss the formal properties of the adaptation operators with respect to this translation, a first question to address is whether the current translation from ARG states to DEOL axioms is faithful. That is, whether this translation is *consequence preserving* and/or *strictly adherent*.

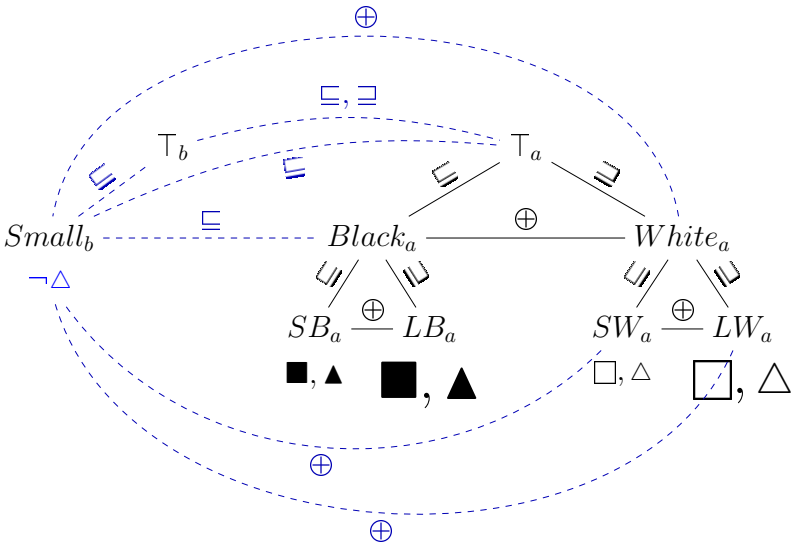

Figure 3.2: Knowledge (black) and belief (blue) of agent a in the translated ARG state of Figure 2.1.

3.2.2 Faithfulness of τ

Faithfulness of the translation τ that maps ARG states to DEOL axioms can be addressed with respect to two questions:

1. Can anything entailed independently from an ontology or an alignment also be entailed by the DEOL axioms as knowledge or beliefs of the agent? (**consequence preservation**)

2. Can anything entailed by the DEOL axioms as knowledge or beliefs of the agent also be entailed independently from an ontology or an alignment? (**strict knowledge/belief adherence**)

Both these properties are defined with respect to entailment as defined in Chapter 2 for ontologies and alignments (Definitions 2.4 and 2.8 on pages 14, 17) and Section 3.1 for DEOL (Definition 3.4 on page 49).

Consequence preservation

If the translation is consequence preserving, this means that anything that can be entailed independently from an ontology of alignment of an agents, can also be entailed by the DEOL axioms as knowledge of beliefs of that agent.

In particular, this means that $\mathcal{O}_a \models \phi$ implies $\tau(s) \models K_a\phi$ and $A_{ab} \models_a \gamma$ implies $\tau(s) \models B_a\gamma$. Consequence preservation is formulated in this way, in two equations and agent by agent, because this is the way that the ARG agents use information: they never consider several alignments at once or the ontologies of other agents. No global reasoning is possible to them.

Proposition 3.1 shows that τ indeed is consequence preserving. The argument for this can be split in two cases: (1) whenever, for a given ARG state s, $\tau(s)$ has no model, trivially anything is known or believed by all agents, and (2) whenever $\tau(s)$ does have a model, DEOL uses propositions from a Description Logic language that are interpreted by an interpretation that works equivalently to that of ontologies and alignments. The proof can be found in Appendix A on page 165.

Proposition 3.1 (Consequence preservation)**.** Let s be an ARG state for a set of agents $\mathcal{A}$, then, for each agent $a \in \mathcal{A}$,

$$\forall\phi : \text{ if } \mathcal{O}_a \models \phi \text{ then } \tau(s) \models K_a\phi \tag{3.8}$$

and

$$\forall\gamma : \text{ if } A_{ab} \models_a \gamma \text{ then } \tau(s) \models B_a\gamma \tag{3.9}$$

where the left-hand $\models$ concerns entailment by ontologies and alignments, where $\models_a$ is entailment restricted to agent a (Definitions 2.4 and 2.8 on pages 14, 17), and the right-hand $\models$ concerns entailment in DEOL (Definition 3.4 on page 49).

Thus, the translation preserves, modulo modalities, the information that agents have. Their knowledge representations, as defined by ontologies, are preserved as knowledge in DEOL and their alignments are preserved as beliefs. Given the structure of the translation, it does not seem to introduce arbitrary information. But do the reverse of these statements, *strict adherence*, hold?

Strict knowledge adherence

What about the reverse of Equation 3.8 that can be called *strict knowledge adherence*? Whenever the logical translation $\tau(s)$ of an ARG state s has a model, it is natural to think that this should hold. Indeed, everything that can be deduced by the knowledge of agents must also be deducible from the knowledge representation, the ontology, of the agent. This is because everything the agents know in $\tau(s)$ must have come from the translation of their ontologies. Therefore, if $\tau(s)$ is consistent (so that it has a model), then the translation is strictly knowledge adherent.

However, there may be situations in which $\tau(s)$ does not have a model, but the ontologies are consistent. In such situations, anything can trivially be deduced by $\tau(s)$, including $K_a\bot$ for each agent $a \in \mathcal{A}$. Yet, since the ontologies are consistent, $\mathcal{O}_a \not\models \bot$ for every $a \in \mathcal{A}$. Therefore, the reverse of Equation 3.8 only holds for certain cases.

Let us investigate in what situations $\tau(s)$ does not have a model, but the ontologies are consistent. This occurs, in particular, when the ARG state s is locally inconsistent. Recall that an ontology is locally consistent for a set of alignments if it has an extended model satisfying all correspondences of the alignments (Definition 2.8 on page 17). Therefore, it is locally inconsistent if any extended model of the ontology does not satisfy some correspondence of some alignment.

Proposition 3.2 (Local consistency preservation)**.** Let s be an ARG state for a set of agents $\mathcal{A}$. Then $\tau(s)$ has a model if and only if s is locally consistent.

The proof of Proposition 3.2 can be found in Appendix A on page 166. It uses the *standard DEOL model* for an ARG state.

Definition 3.6 (Standard DEOL Models)**.** Let s be an ARG state for a set of agents $\mathcal{A}$. The *standard DEOL models* $\langle \mathcal{M}^s, w^s \rangle$ for s are defined by letting $\mathcal{M}^s = \langle W^s, (\geqslant_a^s)_{a\in\mathcal{A}}, \Delta^s, I^s \rangle$ and:

- $W^s = \{w^s\} \cup \{w_a\}_{a\in\mathcal{A}}$;
- $\geqslant_a^s = \{\langle w, w\rangle\}_{w\in W^s} \cup \{\langle w^s, w_a\rangle\}$;
- $\Delta^s = \mathcal{D}$;

And the interpretation I^s satisfies:

1. for each world v and each object name $o \in \mathcal{D}$, the object o is interpreted as itself, i.e. $I^s_v(o) = o$,
2. for w^s and each class name $C \in \mathcal{C}_a$, the class C is interpreted following the *standard interpretation* (Definition 2.5 on page 14) $\hat{I}_a$ of $\mathcal{O}_a$,, i.e. $C^{I^s_{w^s}} = C^{\hat{I}_a}$, and
3. for each w_a an interpretation I'_a is chosen such that $I_a \subseteq I'_a$ for $I_a \models_a \bigcup_{b\in\mathcal{A}\setminus\{a\}} A_{ab}$.

Essentially, the way to construct the standard DEOL models $\langle \mathcal{M}^s, w^s \rangle$ is to create a world that satisfies all the ontologies (this is well-defined because

the classes of the ontologies are disjoint), and to create for each agent a more plausible world that satisfies her own ontology and alignments. The latter is achieved through, for each agent $a \in \mathcal{A}$, choosing the interpretation at this more plausible world to be an interpretation of $\mathcal{O}_a$ locally satisfying all the alignments of a. Then indeed, each agent knows her own ontology (it is true in all her accessible worlds, e.g. in the two worlds on top of Figure 3.3 for agent a) and believes her alignments (they are true in the most plausible world, e.g. in the world on top right of Figure 3.3 for agent a). Figure 3.3 shows how to build a standard DEOL model $\langle \mathcal{M}^s, w^s \rangle$ for a locally consistent ARG state s with three agents.

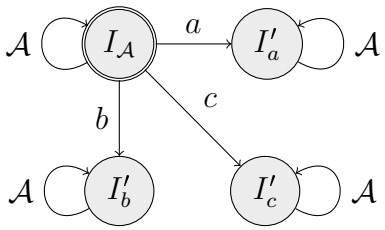

Figure 3.3: A model of the ARG state with three agents, $\mathcal{A} = \{a, b, c\}$. We write $I_{\mathcal{A}}$ as abbreviation for the interpretation that assigns to classes $C \in \mathcal{C}_a$ and objects of $\mathcal{D}$ the standard interpretation I_a (Definition 2.5 on page 14) for agent a.

The standard DEOL models of s are indeed models of $\tau(s)$.

Lemma 3.1. For all ARG states s, the standard DEOL model $\langle \mathcal{M}^s, w^s \rangle$ is a model of $\tau(s)$.

The proof follows from the construction of $\langle \mathcal{M}^s, w^s \rangle$ and can be found in Appendix A on page 166.

Proposition 3.2 shows that local consistency is preserved from ARG to DEOL and vice versa. It uses a procedure to construct a standard DEOL model of a locally consistent ARG state s such that it satisfies $\tau(s)$ (Definition 3.6 on page 55).

Local consistency does not imply global consistency. Consider Example 3.1.

Example 3.1. Let s be the ARG state depicted in Figure 3.4 with three agents a, b and c and their individual ontologies and alignments. It is clear that the alignments are locally consistent, but not globally: combining the alignments we arrive at $C' \sqsupseteq A'$ (via alignments A_{ab} and A_{bc}) and $C \sqsupseteq A'$ (via

the alignment A_{ac}), whereas $C \oplus C'$ (in ontology $\mathcal{O}_c$) and none of these classes can be assigned the empty set (Definition 2.2 on page 12). However, since the alignments are locally consistent, there exists a model by Proposition 3.2. And indeed, this is true because the alignments are private and hence each agent only has access to the alignments in which that agent is involved. To illustrate this, agent a only has access to the blue and red alignments, but not green, in Figure 3.4, agent b to the blue and green ones and agent c to the red and green ones.

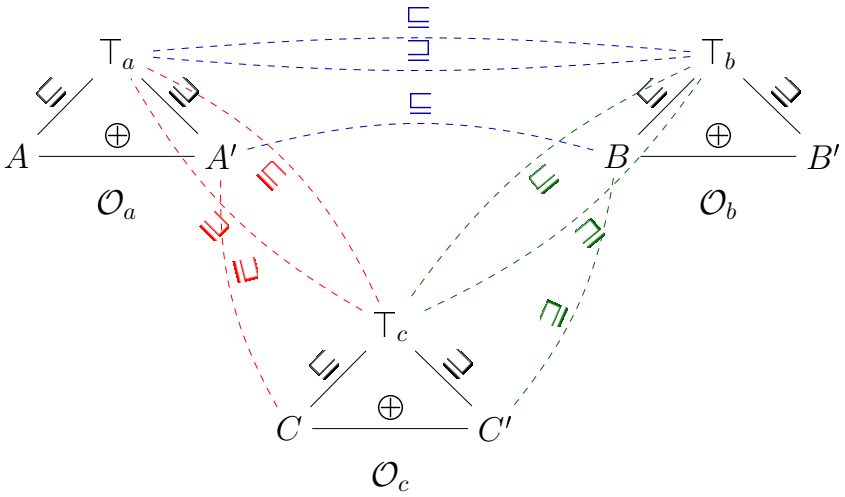

Figure 3.4: The ontologies (black) of agent a (left), agent b (right) and agent c (below) and the alignments (blue, dashed) between them that is globally inconsistent.

Proposition 3.2 shows that global consistency is not required for $\tau(s)$ to have a model. In that sense, DEOL is very faithful to the ARG agents who do not perform any global reasoning.

Whether the reverse of Equation 3.8 holds depends on whether $\tau(s)$ has a model. Therefore, on whether s is locally consistent. The reverse of Equation 3.8 is called *strict knowledge adherence*: all the knowledge of an agent in $\tau(s)$ is already entailed by her ontology in s.

Proposition 3.3 (Strict knowledge adherence)**.** Let s be a *locally consistent* ARG state for a set of agents $\mathcal{A}$, then, for each agent $a \in \mathcal{A}$,

$$\forall \phi : \text{if } \tau(s) \models K_a \phi \text{ then } \mathcal{O}_a \models \phi \tag{3.10}$$

where the first $\models$ concerns entailment by DEOL and the second $\models$ concerns entailment by ontologies.

The proof of Proposition 3.3 can be found in Appendix A on page 167.

Strict belief adherence

Yet, the reverse of Equation 3.9 for beliefs and alignments, called *strict belief adherence*, does not hold. This is because, in DEOL, agents may combine beliefs that are acquired from different alignments to arrive at a new belief of a correspondence between two classes that both belong to other agents' ontologies and not their own. Furthermore, agents may combine their knowledge and beliefs to reach new beliefs, and even anything that is known by an agent is also believed by her.

On ARG, however, agents consider alignments one by one and do not combine their ontologies and alignments. Hence some beliefs entailed by the translation may not be entailed by the alignment. Consider Example 3.2.

Example 3.2. Let s be the ARG state in Figure 3.4 with

$$\begin{aligned} A_{ab} = A_{ba} &= \{\langle B, A', \sqsupseteq\rangle, \langle \top_b, \top_a, \sqsupseteq\rangle\} \\ A_{bc} = A_{cb} &= \{\langle C', B, \sqsupseteq\rangle, \langle \top_c, \top_b, \sqsupseteq\rangle\} \\ A_{ac} = A_{ca} &= \{\langle C, A', \sqsupseteq\rangle, \langle \top_c, \top_a, \sqsupseteq\rangle\} \end{aligned}$$

Then $\tau(s) \models B_b(B \sqsupseteq A')$ and $\tau(s) \models B_b(C' \sqsupseteq B)$, so that $\tau(s) \models B_b(C' \sqsupseteq A')$. But it is clear that $A_{ba} \not\models_b C' \sqsupseteq A'$ nor $A_{cb} \not\models_b C' \sqsupseteq A'$ because $A', C' \notin \mathcal{C}_b$.

Therefore Example 3.2 shows that the reverse of Equation 3.9 does not hold. In particular:

$$\exists \gamma : \tau(s) \models B_a\gamma \wedge A_{ab} \not\models_a \gamma \tag{3.11}$$

Conclusion

The translation from ARG states to DEOL axioms is quite faithful with the difference that agents in DEOL reason globally on agent alignments, whereas ARG agents reason locally. This is present in both the restriction of using locally consistent states and the absence of strict belief adherence

The next step in formally defining and establishing the formal properties of the adaptation operators is to translate the adaptation operators as dynamic upgrades, α in Figure 3.1.

3.2.3 Adaptation Operators as Dynamic Upgrades (δ)

During the gameplay of ARG, agents communicate with each other and through their communication, they learn new information. From the standpoint of DEOL, there are two dynamic acts involved: the communication of $C_b(o)$ in step 2 of ARG and the adaptation operator applied in step 3 of ARG (Definition 2.12 on page 20). With a formal model of ARG, i.e. the translation of ARG states as DEOL axioms, it remains to translate the communication taking place in ARG to DEOL. Then, it can be studied how the knowledge and beliefs of the agents evolve under this communication and whether the adaptation operators as defined in [46, 49] are sufficient and complete to account for these changes.

Recall the diagram shown in Figure 3.1 in which both a translation from ARG states to DEOL axioms and from adaptation operators to dynamic upgrades are depicted.

$$\begin{array}{ccc} \text{ARG state } (s) & \xrightarrow{\quad\tau\quad} & \text{DEOL theory } (\tau(s)) \\ \Big\downarrow \alpha & \overset{\delta}{\vdash\!\!\dashrightarrow} & \delta(\alpha) \Big\downarrow \\ \text{ARG state } (\alpha(s)) & \xrightarrow{\quad\tau\quad} & \text{DEOL theory } (\tau(\alpha(s))) \end{array}$$

Figure 3.1: The translation from ARG states (s) to DEOL theories (τ) and from adaptation operators (α) to dynamic upgrades (δ). (repeated from page 51)

In this section, δ is formally defined. In particular, the communication taking place in ARG, among which the adaptation operators, are translated to dynamic upgrades in DEOL, i.e. announcements and conservative upgrades. One round of ARG, as defined in Definition 2.12 (page 20) with adaptation operator α applied to correspondence C_aRC_b and object o, is a theory transformation that is defined as follows.

Definition 3.7 (ARG Dynamics in DEOL)**.** Let $T = \tau(s)$ be the DEOL theory that is the translation of an ARG state s, let α be an adaptation operator and $\langle C_a, C_b, \sqsupseteq\rangle$ be the correspondence used for object o. Then $\delta(\alpha[\langle C_a, C_b, \sqsupseteq\rangle, o]) : T \to T'$ is a theory transformation, where T' is defined as:

$$T' := \begin{cases} T \cup \{\langle !C_b(o)\rangle \top\} & \text{if } \mathcal{O}_a \vDash C_a(o) \\ T \cup \{\langle !C_b(o)\rangle \top,\ \big\langle d(\alpha[\langle C_a, C_b, \sqsupseteq\rangle, o])\big\rangle \top\} & \text{if } \mathcal{O}_a \nvDash C_a(o) \end{cases}$$

and $d(\alpha[\langle C_a, C_b, \sqsupseteq\rangle, o])$ is a conservative upgrade of the correspondence deleted or added by the adaptation operator (Definition 3.8 on page 61). In the following, also $d(\alpha)$ is written for $d(\alpha[\langle C_a, C_b, \sqsupseteq\rangle, o])$ whenever the correspondence and object are clear from the context.

In the following we will also write $T^{\delta(\alpha)}$ for the $\delta(\alpha)(T)$ or $T^{!C_b(o)}$, in case of a success, and $T^{!C_b(o);d(\alpha[\langle C_a,C_b,\sqsupseteq\rangle,o])}$, or simply $T^{!C_b(o);d(\alpha)}$, in case of a failure. Similarly, we write $d(\alpha)$ for $d(\alpha[\langle C_a, C_b, \sqsupseteq\rangle, o])$ whenever the correspondence $\langle C_a, C_b, \sqsupseteq\rangle$ and object o are clear from the context.

The announcements and conservative upgrades are added in the diamond-form ($\langle\dagger\phi\rangle\psi$) because of the intended meaning: 'ϕ and, after announcing ϕ, ψ' (Section 2.2 on page 26). This ensures that the communication between agents is added to the theory, because if it would appear in box-form, it would be a tautology.

The choice of the translation for the adaptation operators as logical dynamics is quite natural: it is based on the order of dynamic acts on ARG and the trustworthiness of the statements. In ARG, communication of the class occurs first, followed by a belief revision when the adaptation operator is applied.

$\langle!C_b(o)\rangle\top$ First, the statement $C_b(o)$ is communicated in step 2 of the game (Definition 2.12 on page 20) by agent b. This statement is known to agent b because this is part of her ontology and therefore modeled as an announcement. Announcing $C_b(o)$ is performed in complete confidence – it is hard information, a fact – and therefore agent a has no reason to doubt about it.

$\langle d(\alpha)\rangle\top$ Second, the adaptation operator $d(\alpha[\langle C_a, C_b, \sqsupseteq\rangle, o])$ is applied as a belief revision strategy. It specifies how to revise the alignment, which the agents believe to be true, upon a communication failure. But compared to the statement $C_b(o)$, the correspondence added by the adaptation operator is not necessarily correct and may prove to be incorrect at a later stage of the game upon reaching a new communication failure. Because $d(\alpha)$ acts as a belief revision strategy, the DEOL upgrade for the correspondence added is not an announcement but a conservative upgrade – it is soft information, a belief.

Recall the effect of the adaptation operators illustrated in Figure 2.3.

The definitions of the different adaptation operators (Definition 2.13 on page 21) motivate the following definition of adaptation operators as dynamic upgrades.

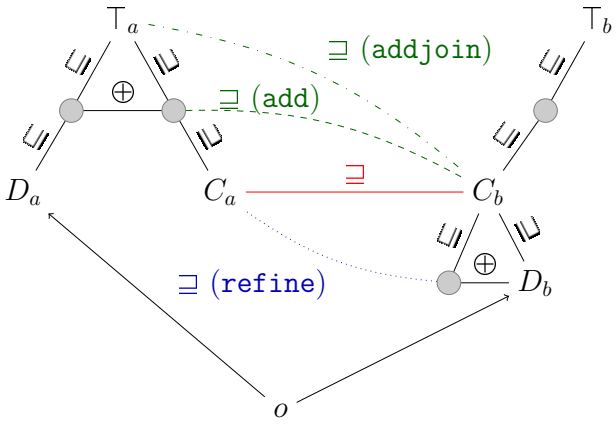

Figure 2.3: Schematic diagram of the adaptation operators for ARG. (repeated from page 23)

Definition 3.8 (Adaptation Operators as Dynamic Upgrades)**.** Let a correspondence $\langle C_a, C_b, \sqsupseteq\rangle$, an object o, and an adaptation operator α be given, then $d(\alpha)$ is defined as:

$$
\begin{aligned}
d(\texttt{delete}) &= \uparrow(\neg(C_a \sqsupseteq C_b)) \\
d(\texttt{add}) &= \uparrow(\neg(C_a \sqsupseteq C_b) \ \wedge\ msc_a(C_a) \sqsupseteq C_b) \\
d(\texttt{addjoin}) &= \uparrow(\neg(C_a \sqsupseteq C_b) \ \wedge\ mscc_a(o, C_a) \sqsupseteq C_b) \\
d(\texttt{refine}) &= \uparrow(\neg(C_a \sqsupseteq C_b) \ \wedge \bigwedge_{C'_b \in M_b(C_b,o)} C_a \sqsupseteq C'_b) \\
d(\texttt{refadd}) &= \uparrow(\neg(C_a \sqsupseteq C_b) \ \wedge\ mscc_a(o, C_a) \sqsupseteq C_b \wedge \bigwedge_{C'_b \in M_b(C_b,o)} C_a \sqsupseteq C'_b)
\end{aligned}
$$

where $M_b(C_b, o) = \{C'_b \in mgcx_b(C_b, o) \mid \nexists\langle C'_a, C'_b, \sqsupseteq\rangle \in A_{ab}\}$.

We use conservative upgrades and not radical upgrades. In fact, as long as we do not consider a notion of "strong belief", there is no difference between the two concerning the beliefs revised by the agents (see the discussion in Section 2.2.1 on page 31). Finally, conservative upgrades have been chosen because they are more general (weaker) than radical upgrades.

Again, as was the case for the adaptation operators on ARG states [46, 49], the translated adaptation operators $d(\texttt{add})$, $d(\texttt{addjoin})$, $d(\texttt{refine})$ and $d(\texttt{refadd})$ entail $d(\texttt{delete})$, and the translated adaptation operator $d(\texttt{refadd})$ entails $d(\texttt{addjoin})$ and $d(\texttt{refine})$. This can be observed directly from Definition 3.8.

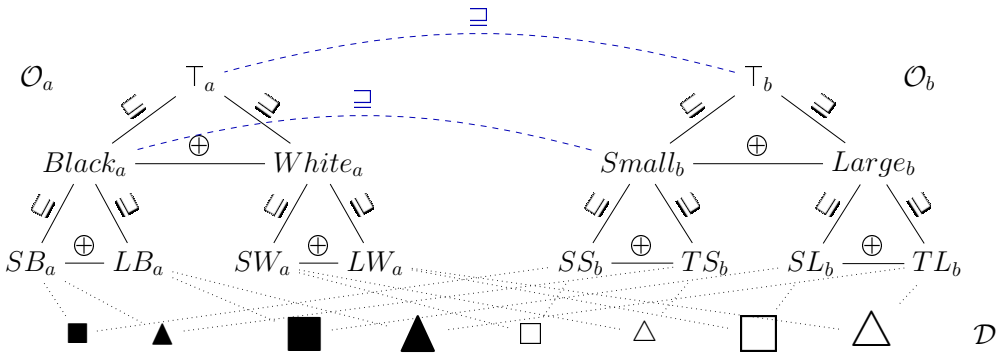

Figure 2.1: The ARG state $s = \langle\{\mathcal{O}_a, \mathcal{O}_b\}, \{A_{ab}\}\rangle$. (repeated from page 13)

The drawback of modeling the operators as announcements and conservative upgrades is that they are model transformers and that therefore, when the ARG state has no model in DEOL, the transformation yields no model. On the contrary, in ARG, agents blindly apply operators without bothering about local or global inconsistency. Experiments went as far as showing that an improved version of ARG (with relaxation) brings agents to a fully consistent state even when starting with an inconsistent one [49]. This is not possible in the DEOL translation because it is restricted to consistent states.

Example

To illustrate the effect of $\delta(\alpha)$, consider two examples concerning the ARG state in Figure 2.1, one in which an object is drawn that leads to a success, and the other in which it is a failure.

Example 3.3 (Success)**.** When ARG is played with ▲, agent b announces that $!Small_b(▲)$ and the correspondence used is $\langle Black_a, Small_b, \sqsupseteq\rangle \in A_{ab}$. This information is compatible with the information of agent a: $Black_a$ is compatible with SB_a, i.e. the most specific class of ▲.

Compared to ARG where the round is now finished, there are additional epistemic-doxastic changes on the corresponding DEOL model. The announcement carries more information than just indicating that the round of ARG was a success, it provides agent a with new knowledge: $K_a(Small_b(▲))$. In other words, agent a is now given concrete evidence that ▲ is a member of $Small_b$. Figure 3.2 can be compared to Figure 3.5 for an overview of the changes to the epistemic-doxastic state of agent a.

Example 3.4 (Failure)**.** If instead ARG is played with △, the round is a failure. Agent b announces $!Small_b(△)$ using the same correspondence

$\langle Black_a, Small_b, \sqsupseteq \rangle \in A_{ab}$. However, this information contradicts the knowledge of agent a and, as a result, the correspondence (belief) of the alignment will be dropped.

However, this is not the only revised belief. The contradicted initial beliefs turn into knowledge of their negation. For example, $B_a(\neg Small_b(\triangle))$ becomes $K_a(Small_b(\triangle))$ after the announcement. Compare also Figure 3.2 and Figure 3.6 for an overview of the changes to the epistemic-doxastic state of agent a.

According to ARG, an adaptation operator is applied adding correspondence(s) to the alignment (Definition 3.8 on page 61). This added correspondence(s) should be amongst the beliefs of the agents at the end of the round of ARG. However, for some operators, the correspondences were *already* believed by agent a before the adaptation operator is applied.

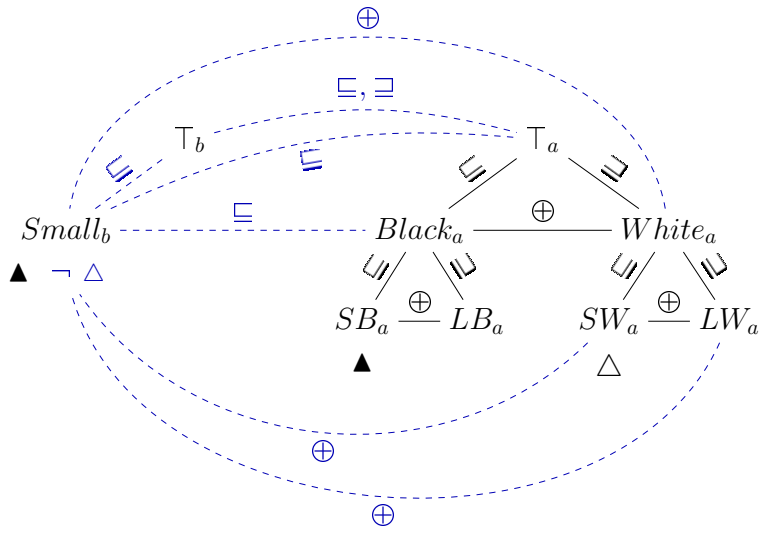

Figure 3.5: The knowledge (solid black) and belief (dashed blue) of agent a of Example 2.2 after the announcement $!Small_b(\blacktriangle)$.

3.3 Formal Properties of the Adaptation Operators

With the faithful translation from ARG to DEOL, a bridge is established between *experimental* cultural knowledge evolution and a *theoretical* model of

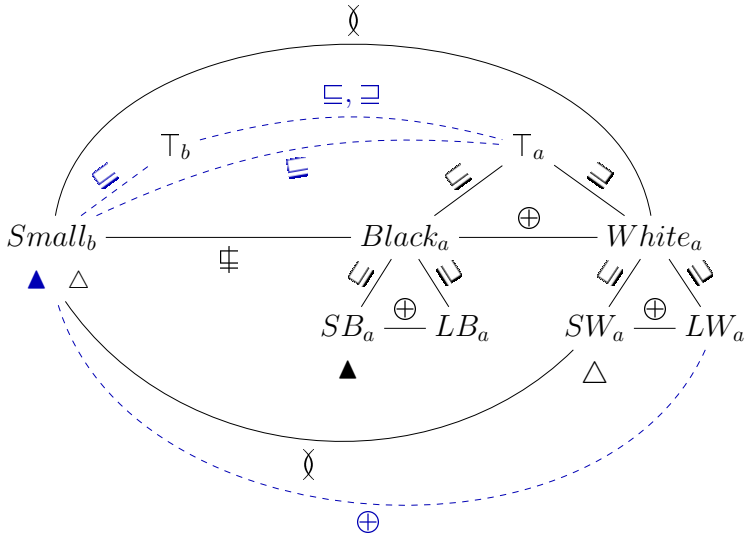

Figure 3.6: The knowledge (solid black) and belief (dashed blue) of agent a of Example 2.2 after the announcement $!Small_b(\triangle)$.

cultural knowledge evolution in Dynamic Epistemic Ontology Logic (DEOL). This bridge is necessary to complement the experimental results about ARG, i.e. that agents improve their alignments and reach successful communication [46, 49], with a theoretical model of the formal properties of the adaptation operators that until now has been lacking.

In this Section, we investigate how the adaptation operators applied in ARG compare to the mechanisms in logic for agents to evolve their knowledge and beliefs. In particular, the formal properties correctness, completeness and redundancy are defined with respect to the translation and it is proven that all but one adaptation operator are formally correct, all adaptation operators are incomplete and most adaptation operators are partially redundant. Finally, these results are discussed.

Recall the diagram in Figure 3.1 (page 51) that will be used to define correctness, completeness and redundancy. For clarity when defining the definitions, the arrows are colored, see Figure 3.7

3.3.1 Correctness

An adaptation operator is correct if applying the translation τ and then the logical dynamics $\delta(\alpha)$ (following the red lines in Figure 3.7) entails applying first the adaptation operator α and then the translation τ (following the blue

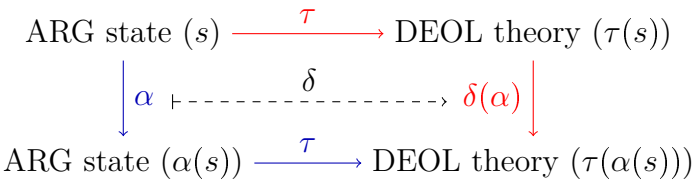

Figure 3.7: The translation from ARG states (s) to DEOL theories (τ) and from adaptation operators (α) to dynamic upgrades (δ). The colored arrows are used in the definitions of correctness and completeness. The red lines take the upper left ARG state s to the upper right DEOL theory $\tau(s)$ and then to the lower right DEOL theory $\tau(\alpha(s))$, and the blue lines take the upper left ARG state s to the lower left ARG state $\alpha(s)$ and then to the lower right DEOL theory $\tau(\alpha(s))$.

lines in Figure 3.7).

Definition 3.9 (Correctness)**.** An adaptation operator α is correct if and only if $\forall s$: $(\tau(s))^{\delta(\alpha)} \models \tau(\alpha(s))$.

Most of the adaptation operators are correct, see Proposition 3.4.

Proposition 3.4 (Correctness)**.** The adaptation operators `delete`, `addjoin`, `refine` and `refadd` are correct.

The proof is done for agent a and adaptation operator `addjoin`. The proof for `delete` now follows because `delete` is entailed by `addjoin`, the proof for `refine` is symmetric and the proof for `refadd` is the combination of `addjoin` and `refine`.

Proof. Because `addjoin` only adds beliefs, it suffices to show that these beliefs are entailed (we note C_a^o for $mscc_a(o, C_a)$): $(\tau(s))^{!C_b(o);\Uparrow(C_a \sharp C_b \wedge C_a^o \sqsupseteq C_b)} \models B_i(C_a \sharp C_b) \wedge B_i(C_a^o \sqsupseteq C_b)$ for $i \in \{a, b\}$. This holds because initially the correspondence is believed, i.e. $\tau(s) \models B_i(C_a \sqsupseteq C_b)$, and the upgrade $!C_b(o); \uparrow(C_a \sharp C_b \wedge C_a^o \sqsupseteq C_b)$ deletes all the worlds from $\tau(s)$ in which $C_b(o)$ is false and then rearranges the remaining worlds such that the '$C_a \sharp C_b \wedge C_a^o \sqsupseteq C_b$'-worlds become more plausible than the '$\neg(C_a \sharp C_b \wedge C_a^o \sqsupseteq C_b)$'-worlds. Because there remain '$C_a \sharp C_b \wedge C_a^o \sqsupseteq C_b$'-worlds accessible for both agents, the belief is enforced. For agent b, this is true because the announcement $!C_b(o)$ does not alter her epistemic-doxastic state (she already knew that $C_b(o)$ as it is in her ontology), and for agent a, because the announcement $!C_b(o)$ deletes the worlds in which $C_a \sqsupseteq C_b$ ($\neg C_a(o)$ holds because the correspondence and announcement caused a failure) or $C_a \equiv C_b$ so that both

$C_a \not\equiv C_b$ or $C_a^o \sqsupseteq C_b$ are unchanged. Therefore the beliefs $B_i(C_a \not\equiv C_b)$ and $B_i(C_a^o \sqsupseteq C_b)$ are enforced for agents $i \in \{a, b\}$. Hence `addjoin` is correct. □

Proposition 3.4 excludes one adaptation operator: `add`. This is because `add` is not correct. More precisely, it is not *always* correct. The reason is that `add` does not take into account whether the most specific superclass of C_a, to which a correspondence is added, is consistent with the drawn object o. If it is not, this means that the added correspondence is incorrect and will cause another failure in future rounds. Moreover, if it is consistent, `add` is actually equivalent to `addjoin`.

Proposition 3.5 (Incorrectness of `add`)**.** The adaptation operator $\alpha = \texttt{add}$ is incorrect, i.e. $\exists s : (\tau(s))^{\delta(\texttt{add})} \not\models \tau(\texttt{add}(s))$, and $\forall s$ such that $(\tau(s))^{\delta(\texttt{add})} \models \tau(\texttt{add}(s))$: $\texttt{add}(s) = \texttt{addjoin}(s)$.

Proof. We need to prove the existence of an ARG state s where $(\tau(s))^{\delta(\texttt{add})} \not\models \tau(\texttt{add}(s))$ with upgrade $\delta(\texttt{add}) = !C_b(o); \uparrow (C_a \not\equiv C_b \wedge msc_a(C_a) \sqsupseteq C_b)$, object o s.t. $\mathcal{O}_b \models C_b(o)$ and $\langle C_a, C_b, \sqsupseteq \rangle \in A_{ab}$ the failing correspondence. Pick s to be any such ARG state where the most specific superclass $C' = msc_a(C_a)$ of C_a is incompatible with o, i.e. $\mathcal{O}_a \not\models C'(o)$. Then $\tau(s) \models K_a(\neg C'(o))$ and $(\tau(s))^{\delta(\texttt{add})} \models K_a(C_b(o)) \wedge K_a(C' \not\equiv C_b)$. This is because $\delta(\texttt{add})$ deletes all '$\neg C_b(o)$'-worlds from $\tau(s)$ and therefore also all the worlds accessible by agent a where $C \sqsupseteq C_b$ for C such that $\tau(s) \models K_a(C(o))$. In particular, this holds for $C' = msc_a(C_a)$. But, after applying the adaptation operator `add`, $\langle C', C_b, \sqsupseteq \rangle$ becomes part of the alignment, so that $\tau(\texttt{add}(s)) \models B_a(C' \sqsupseteq C_b)$. Hence $(\tau(s))^{\delta(\texttt{add})} \not\models \tau(\texttt{add}(s))$.

Moreover, whenever $(\tau(s))^{\delta(\texttt{add})} \models \tau(\texttt{add}(s))$ it must be that $\mathcal{O}_a \models C'(o)$ so that, per definition, $C' = msc_a(C_a) = mscc_a(o, C_a)$, i.e. `add` is equivalent to `addjoin`. □

It must be noted that because `add` may be incorrect, the repair may not be satisfactory immediately, but will be in the long run as shown by the experiments in [46, 49].

Proposition 3.5 is in line with initial predictions and experimental results [49]: `addjoin` shows faster convergence to the same result than `add`. This is because `add` can force false correspondences to be added to the alignment that can later cause a failure. From the results presented here, this prediction is established and it is clear that for a logical agent, `add` should be abandoned.

3.3.2 Redundancy

After establishing correctness, let us look at another formal property of the adaptation operators: redundancy. In Example 3.4, some redundancy by the

adaptation operators occurred: some of the correspondences added by the different adaptation operators were already believed by the agents after the announcement of $C_b(o)$. It turns out that this is not a coincidence.

For logical agents, some adaptation operators are redundant for *every* ARG state but only for one of the agents: `delete` and `addjoin` are redundant with respect to agent a.. This is called *partial redundancy*. Before partial redundancy is formally defined, first *redundancy* is defined. An adaptation operator α is redundant if and only if applying $!C_b(o)$ is already sufficient to cause the beliefs of the agents to be revised in line with α. More specifically, if applying $!C_b(o)$ to the DEOL translation of s, i.e. $\tau(s)$, leads to an interpretation of the DEOL translation of $\alpha(s)$, i.e. $\tau(\alpha(s))$.

Definition 3.10 (Redundancy)**.** An adaptation operator α is *redundant* if and only if $\forall s$: $(\tau(s))^{!C_b(o)} \models \tau(\alpha(s))$.

See Figure 3.8 for an illustration of an adaptation operator α that is redundant.

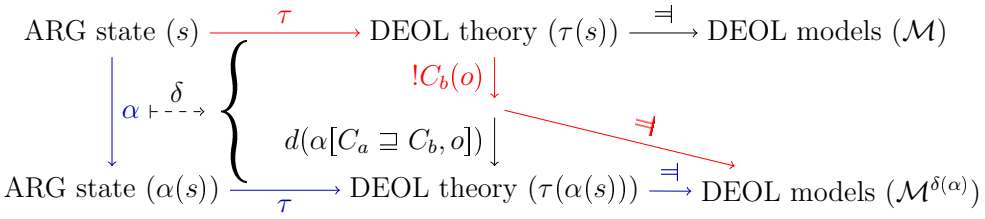

Figure 3.8: The translation from ARG states (s) to DEOL theories (τ), interpreted in DEOL models $(\mathcal{M})$, and from adaptation operators (α) to dynamic upgrades (δ) for redundant adaptation operators. Hence, $\mathcal{M}^{\delta(\alpha)}$ is a model of both $\tau(\alpha(s))$ (following the blue lines) as well as $\tau(s)^{!C_b(o)}$ (following the red lines).

When the announcement $!C_b(o)$ is made, agent a compares the class C_a such that $C_a \sqsupseteq C_b \in A_{ab}$ (i.e. her belief $B_a(C_a \sqsupseteq C_b)$) and the classifications of object o (i.e. her knowledge $D(o)$ for all D such that o belongs to it), but agent b cannot do this as she does not know the class of o in agent a's ontology. Therefore, even though for agent a her knowledge and beliefs may become contradictory (in case $K_a(\neg C_a(o))$, leading her to drop her belief of the correspondence, agent b does not have such a mechanism to drop $C_a \sqsupseteq C_b$. Therefore, `delete` is not redundant and because every adaptation operator extends `delete`, this holds for all adaptation operators.

Proposition 3.6 (No redundancy)**.** No adaptation operator is redundant.

Proof. Let s be an ARG state and $\langle C_a, C_b, \sqsupseteq \rangle \in A_{ab}$ be the failing correspondence with object o. Because agent b already knows $C_b(o)$, the announcement $!C_b(o)$ does not alter her knowledge and beliefs, in particular her belief of the correspondence, i.e. $(\tau(s))^{!C_b(o)} \models B_b(C_a \sqsupseteq C_b)$. However, after applying the adaptation operator α to s, whatever the adaptation operator, the correspondence $\langle C_a, C_b, \sqsupseteq \rangle$ is deleted from A_{ab}. I.e. $\tau(\alpha(s)) \not\models B_b(C_a \sqsupseteq C_b)$. Hence $(\tau(s))^{!C_b(o)} \not\models \tau(\alpha(s))$ for any adaptation operator α and thus no adaptation operator is redundant. □

The adaptation operators discussed here are not redundant, but sometimes *partially redundant*. This means that they may be redundant with respect to one agent, but not with respect to both. This agent must be agent a, because even `delete` is not redundant for agent b, as was shown in the proof of Proposition 3.6. To prove partial redundancy for agent a and adaptation operator α, we show that whatever is known and believed by a in $\tau(\alpha(s))$ is also known and believed by a in $\tau(s)^{!C_b(o)}$. In fact, because adaptation operators only alter the beliefs of agents, it suffices to show partial redundancy by showing that this holds for beliefs. That means that $!C_b(o)$ is enough for agent a to revise her beliefs in accordance to the adaptation α.

Definition 3.11 (Partial Redundancy)**.** An adaptation operator α is *partially redundant* for agent a if and only if $\tau(\alpha(s)) \models B_a\phi$ implies $(\tau(s))^{!C_b(o)} \models B_a\phi$ for any ARG state s and any ϕ in $\mathcal{L}_{DEOL}$.

Proposition 3.7 (Partial Redundancy)**.** The adaptation operators `delete` and `addjoin` are partially redundant with respect to agent a.

The proof is done for the adaptation operator `addjoin`. The proof for `delete` now follows because it is entailed by `addjoin`.

Proof. Suppose that $\tau(\texttt{addjoin}[\langle C_a, C_b, \sqsupseteq \rangle, o](s)) \models B_a\phi$. Since the adaptation operators can be considered as contraction operators (see page 23), compared to $\tau(s)$, the correspondences $C_a \sqsupseteq C_b$ and $msc_a(C_a) \sqsupseteq C_b$ and what follows from them in combination with the agents' ontology by logical closure are deleted from the beliefs of agent a in the translated state of `addjoin` (with correspondence $\langle C_a, C_b, \sqsupseteq \rangle$ and object o) applied to s. Therefore it must be that $\phi \in \{\psi \mid \tau(s) \models B_a\psi\} \backslash Cl_{\mathcal{O}_a}(\{\phi \notin \{C_a \sqsupseteq C_b, msc_a(C_a) \sqsupseteq C_b\})$, where $Cl_{\mathcal{O}_a}(X)$ of a set X is the logical closure of X with respect to $\mathcal{O}_a$, i.e. the set of those formulas that can be deduced from the combination of X and $\mathcal{O}_a$ but not by X or $\mathcal{O}_a$ alone. We have to show that for these ϕ also $\tau(s)^{!C_b(o)} \models B_a\phi$. But this already holds because the only beliefs deleted by the announcement $!C_b(o)$ are also those in $Cl_a(\{C_a \sqsupseteq C_b, msc_a(C_a) \sqsupseteq C_b\}$ because they are

incompatible with $C_b(o)$ and $\mathcal{O}_a$ (because $\mathcal{O}_a \models \neg C_a(o) \wedge \neg(msc_a(C_a))(o)$), see Figure 3.9. Hence, `addjoin` is partially redundant for agent a.

□

In Figure 3.9 the knowledge and beliefs of agent a are illustrated before and after the announcement $!C_b(o)$ for an intuition.

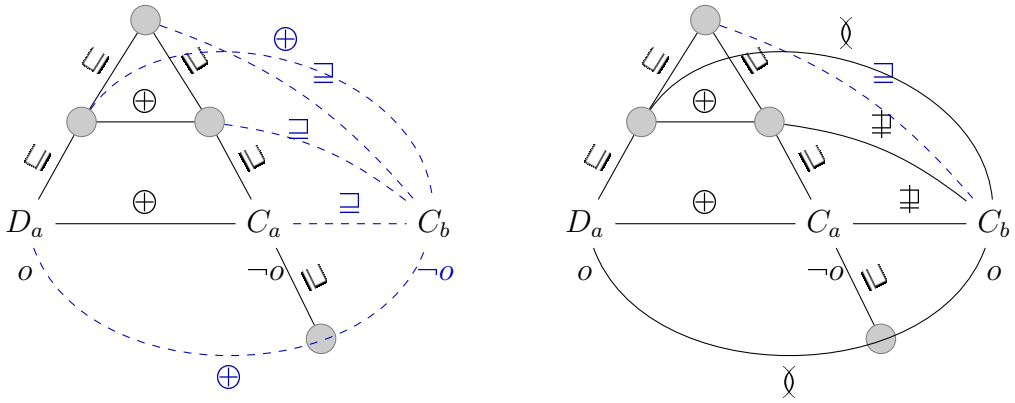

Figure 3.9: The knowledge (solid black) and beliefs (dashed blue) of agent a before (left) and after (right) the announcement $!C_b(o)$.

The fact that none of the adaptation operators are redundant with respect to both agents shows that they cannot be fully discarded. Even the simple `delete` carries valuable information to agent b: namely that the initial correspondence fails. Without this operator, agent b would not be aware whether the round of ARG is a success or a failure.

3.3.3 Incompleteness

Finally, the formal property completeness of the adaptation operators is considered: do the operators capture all the information that can be learned, compared to logical agents? Intuitively, this is proved by comparing what is learned by the agents in ARG scenarios from application of the adaptation operators with what is learned by logical agents in DEOL from the dynamic upgrades. If the former implies the later, the operator is (epistemically) complete.

Recall again the diagram in Figure 3.7.

To show that the adaptation operators are complete, we must show that this diagram commutes in the reverse direction compared to correctness. More precisely, that, for a given ARG state s, applying first the adaptation

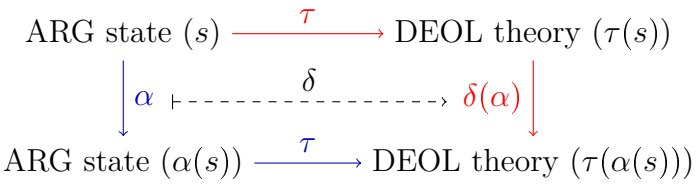

Figure 3.7: The translation from ARG states (s) to DEOL theories (τ) and from adaptation operators (α) to dynamic upgrades (δ). The colored arrows are used in the definitions of correctness and completeness. (repeated from page 65)

operator α and then the translation τ (following the blue lines above) entails applying the translation τ and then the logical dynamics $\delta(\alpha)$ (following the red lines above).

Definition 3.12 (Completeness)**.** An adaptation operator α is complete if and only if $\forall s$: $\tau(\alpha(s)) \models (\tau(s))^{\delta(\alpha)}$.

All adaptation operators are incomplete, see Proposition 3.8.

Proposition 3.8 (Incompleteness)**.** All adaptation operators are incomplete.

Proof. Again, consider the knowledge and beliefs of agent a before and after the announcement $!C_b(o)$, see also Figure 3.9. After the announcement $!C_b(o)$, agent a receives concrete information that object o belongs to the class C_b, i.e. she comes to *know* this information: $(\tau(s))^{!C_b(o)} \models K_a(C_b(o))$. And, by definition, this knowledge remains after application of *any* conservative upgrade, i.e. $(\tau(s))^{\delta(\alpha)} \models K_a(C_b(o))$. Yet, this knowledge is never acquired through application of the adaptation operators because they only concern the alignment, i.e. beliefs of class relations, and not knowledge of instance classification. Hence $\tau(\alpha(s)) \not\models K_a(C_b(o))$ and $\tau(\alpha(s)) \not\models (\tau(s))^{\delta(\alpha)}$. □

This means that, even though the adaptation operators are not redundant, and therefore carry valuable information for at least one agent, the agents do not take full advantage of what could be learned.

The proof for incompleteness uses the announcement $!C_b(o)$ that causes agent a to come to know $C_b(o)$. In fact, *all* agents $a \in \mathcal{A}$ come to know this. Switching to a private announcement $!_{\{a\}}C_b(o)$ avoids this, but incompleteness will still hold because also through $!_{\{a\}}C_b(o)$, agent a will come to know $C_b(o)$, which is discarded on ARG.

3.4 Discussion

In this chapter, it is proved that most adaptation operators are correct, all are incomplete and some are partially redundant. This benefits the theoretical understanding of the mechanisms agents use to evolve their knowledge and complement the experimental results [46, 49].

However, despite of the lack of formal properties, experiments have shown that ARG works quite well in practice: through application of the adaptation operators, agents can improve their alignments and reach successful communication [46, 49]. Furthermore, the adaptation operators are ARG state preserving (Property 2.1 on page 24), produce safe and entailed alignments [49], and work even in the case of globally inconsistent networks of ontologies and alignments (Figure 3.4 on page 57). Why is that, despite the lack of formal properties? In short, this is because agents do not need to be logical to communicate successfully.

For example, the proof of incompleteness of the adaptation operators is based on the fact that the adaptive agents in ARG do not memorize the objects that led to failures in communication and to apply an adaptation operator to revise the alignment. In particular, they forget to which class the other agent classifies the object, but this classification is deduced knowledge in DEOL through the announcement $!C_b(o)$. Nevertheless, even without this knowledge, agents can communicate successfully. They focus on the alignment and on how to improve the alignment through communication. This also means that, despite forgetting the class of the object, ARG agents can communicate successfully. That is, they do not need the concrete examples (the object that caused the failure of the correspondence) in order to communicate with each other, as long as they can use more general conclusions (the improved alignment). Hence, they do not need to remember the cases that led them to these conclusions.

This discussion may also be linked to the proof of redundancy: it is based on the fact that the same announcement $!C_b(o)$ provides extra knowledge to logical agents not available to ARG agents. One may think that the announcement, playing an important role in the incompleteness and redundancy proofs of the adaptation operators, could have been avoided in the translation of the game in order to prevent incompleteness and redundancy. Or less rigorous, to replace the public announcement by a private announcement only to agent a. This would avoid *all* agents to learn $C_b(o)$ from the round of ARG between agents a and b. However, the main objection is that the current translation is indeed faithful to the communication taking place in the ARG game: agents announce to which class the object belongs, they use this fact to improve their alignment and then forget it. Furthermore, a

private announcement $!_{\{a\}}C_b(o)$ would still cause incompleteness and redundancy as the proofs are only concerning agents a and b. It can therefore also be argued that what is missing in the translation is a form of forgetting.

There are two ways to interpret our results: (1) either the adaptive agents use sub-logical behavior, (2) or the logical model of ARG in DEOL is insufficient to model them and not faithful to them, although the faithfulness results show that it is very close. Both interpretations are correct and compatible, but addressing them dictates different courses of action.

One direction is to implement agents in ARG that reason more faithfully to the DEOL logic by introducing new adaptation operators that repair the alignment correctly, completely and without redundancy. It may then be expected that the adaptation operators would correspond better to the mechanisms used in logic for agents to evolve their knowledge and beliefs. Although this was not the goal of the initial experiments, providing agents with more logical reasoning power could be considered. However, it must be noted that fully logical agents may not be desirable because they are not the panacea: recall that for locally inconsistent ARG states, the corresponding logical theory has no model (Proposition 3.2 on page 55) so that the agents may deduce anything, whereas ARG can deal with this situation fine. In ARG, a locally inconsistent state does not cause the agents to be lost: they repair the alignment until they reach a consistent state. Hence, such an approach would restrict the adaptive agents, making them unable to repair the alignments whenever it is inconsistent.

Implementing logical behavior for ARG, which in itself is an interesting research question, is also not the goal of this thesis. The goal is to understand cultural knowledge evolution and the mechanisms used in this framework by agents to evolve. Therefore, the other way to interpret the results consists in bringing the logic closer to the agents. To find the bridge between experimental cultural knowledge evolution and a theoretical analysis of cultural knowledge evolution in logic, the differences between adaptive agents and logical agents need to be further explored. It was already mentioned that adaptive agents are unable to remember the individual cases because they focus on general knowledge, whereas logical agents cannot discard these. Other differences are that the adaptive agents reason locally (they use the alignments one by one), whereas the logical agents combine their knowledge and beliefs to reason globally, and that the logical agents share a fixed vocabulary, preventing them from using heterogeneous knowledge representations like adaptive agents.

The next chapter is dedicated to further explore the differences in detail and explain the directions to be taken.

3.5 Conclusion

In order to understand the formal properties of adaptation operators used in the alignment repair game, ARG was translated into a dynamic epistemic logic, DEOL, with embedded ontological statements. The faithfulness of this translation was assessed by showing that it preserves consequences of ontologies and alignments and that the generated knowledge is the same if the game states are consistent. With DEOL, we proved that all but the `add` operator are correct, that `delete`, `addjoin` and `refine` are redundant for one agent, and that all adaptation operators are incomplete. These results complement the experimental ones in theoretically comparing the different adaptation operators.

In spite of, or because of, the simplicity of ARG, modeling ARG in logic revealed more challenging than expected. Developing a tighter connection requires addressing fundamental issues that give rise to the differences between adaptive agents and logical agents that are discussed in the next chapter.

Chapter 4

Fundamental Differences between Adaptive Agents and Logical Agents

> But that's men all over ... Poor dears, they can't help it. They haven't got logical minds.
>
> Dorothy Leigh Sayers

The previous chapter established the formal properties of the adaptation operators. In particular, it was shown that all but the `add` operator are correct, that `delete` and `addjoin` are redundant for one agent, and that all adaptation operators are incomplete. It was then discussed how these results relate to the experimental results that through playing ARG, agents improve their alignments and converge to successful communication [46, 49]. Two explanations were provided: (1) either the agents playing ARG use sub-logical behavior, or (2) the logical model, though faithful, is insufficient to describe their knowledge evolution. In this chapter, the second explanation is further developed.

We develop a tighter connection between ARG and its logical model by analysing the differences between *adaptive agents* (those playing ARG) and *logical agents* (those in the model of ARG in logic). Three such differences are identified with the aim to modify the logical model of ARG that better resembles the behavior of adaptive agents:

1. Adaptive agents reason locally while logical agents reason globally,
2. Logical agents share a fixed vocabulary, preventing them from using

heterogeneous knowledge representations like adaptive agents, and

3. Adaptive agents are unable to remember individual cases because they focus on general knowledge, whereas logical agents cannot discard these.

Addressing these differences shed light on some fundamental issues in the logic used to model the adaptive agents, Dynamic Epistemic Ontology Logic (DEOL). But more generally these issues also occur in Dynamic Epistemic Logic (DEL). This is because the differences between adaptive agents and logical agents do not arise from the fact that DEOL uses propositions from a simple Description Logic language, but lay deeper within the framework of DEL.

In the following, each of the three differences is explained and it is discussed what elements of DE(O)L gives rise to it.

4.1 Local versus Global Reasoning

The first difference between adaptive agents and logical agents concerns the extent to which they use their reasoning capacities. In particular, that the adaptive agents reason *locally*, while the logical agents do so *globally*.

In ARG, the adaptive agents use knowledge representations, formally defined in ontologies, to describe what they know and use alignments to translate terms in their ontologies to terms in other agents ontologies. The alignments are used one by one: when an agent communicates with another agent, she picks the alignment between their ontologies and use it to communicate. This ensures that the agents can understand each other, even though their ontologies use different vocabularies. It also means that, during this communication, the other alignments, between her and other agents' ontologies, are not relevant: they are not useful in the current communication. For example, when agent a and b communicate, the alignment used is A_{ab}. Any other alignment A_{ac} or A_{bc}, where c is different from a and b, does not benefit the communication between them. Using the alignments in this way is what is referred to as reasoning *locally*.

In logic, the situation is different. Logical agents have the ability to combine their knowledge and beliefs to arrive at new conclusions as described by the axiom schemata K for DEL (see Table 2.1 on page 30).

$$\big(B_a\phi \wedge B_a(\phi \to \psi)\big) \to B_a\psi \tag{4.1}$$

And more particularly for DEOL, the axiom schemata 3.3-3.7 (page 50) can be used obtain the following schemata for combining beliefs:

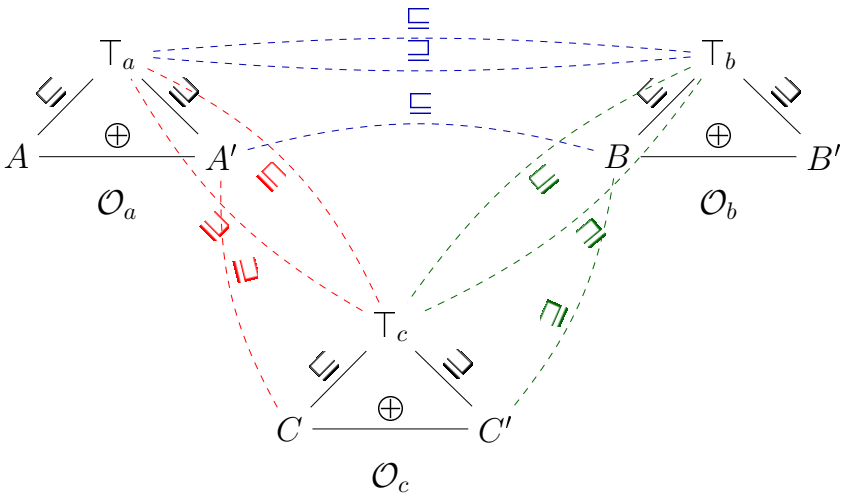

Figure 3.4: The ontologies (black) of agent a (left), agent b (right) and agent c (below) and the alignments (blue, dashed) between them that is locally consistent, but globally inconsistent. (repeated from page 57)

$$\big(B_a(C(x)) \wedge B_a(C \sqsubseteq D)\big) \rightarrow B_a(D(x)) \tag{4.2}$$

$$\big(B_a(C \sqsubseteq C') \wedge B_a(C' \sqsubseteq D)\big) \rightarrow B_a(C \sqsubseteq D) \tag{4.3}$$

Where these schemata also apply whenever one or both of the conjuncts concerns knowledge instead of belief of an agent because in logic, knowledge implies belief [42].

The schemata imply that, when ARG is translated to DEOL and ontologies become agents' knowledge and alignments agents' beliefs, logical agents combine their ontologies and alignments or combine different alignments to deduce new correspondences or classifications as beliefs. Recall the ARG state in Figure 3.4, in which the translation allows agent a to combine her beliefs that $A' \sqsubseteq B$ and $C \sqsubseteq A'$ to arrive at a the belief that $C \sqsubseteq B$. This means that logical agents reason what will be referred to as *globally*: they combine whatever information available to them, both in the form of knowledge and beliefs, and deduce everything that can be deduced from it. In other words, logical agents make use of deductive closure, which is not available to adaptive agents.

Actually, there are other types of global reasoning that are different from the one used by the logical agents in the model of ARG in DEOL. In the logical model, the agents combine all the alignments they have with their

own ontology. However, alignments between other agents' ontologies and ontologies by other agents are not accessible to them. One may also consider that alignments are public or even all ontologies and alignments are. In the second situation, the requirement of agents using heterogeneous knowledge representations is violated and leads to the situation of a universally shared ontology, which comes at the price of autonomy. Public alignments, but private ontologies, on the other hand, lead agents to be able to discover locally consistent, but globally inconsistent states. Consider again the example in Figure 3.4. With public alignments, the agents would learn that $C \sqsubseteq A' \sqsubseteq B \sqsubseteq C' \oplus C$, which is a contradiction because classes are non-empty (Definition 2.2 on page 12). In other words, it is possible for agents to compose alignments to generate new alignments, and even entail inconsistent beliefs. This is stronger than the logical model of ARG in DEOL, where both ontologies and alignments are private. For example, the ARG state in Figure 3.4 does not lead to the logical agents having inconsistent beliefs.

The difference between the local reasoning used by adaptive agents and the global reasoning affects the faithfulness of the translation τ. The translation satisfies strict knowledge adherence (Proposition 3.1 on page 54), i.e. whatever the logical agents know can be deduced from the ontology, but not strict belief adherence (Propositions 3.3 on page 57), i.e. there are beliefs inferred by logical agents that cannot be independently deduced from the alignment. These beliefs inferred by logical agents are specifically the beliefs of correspondences and classifications arising from combining their ontology and alignments.

This leads to question the ability of agents to combine their knowledge and beliefs (Equations 4.1, 4.2 and 4.3). This is also called the problem of *logical omniscience* [61], or the *closure property* of logic: the knowledge and beliefs of agents are closed under logical inference. In logic, this is often a desirable property, however, it cannot directly be applied to the model of ARG because the adaptive agents reason locally in different situations – these situations being different alignments, that are used one by one. Therefore, instead of questioning 'closure under logical inference', it should be questioned whether the knowledge and beliefs of agents should be 'closed under different perspectives' in the logical model of ARG.

Whether agent's reasoning should be closed under different perspectives was already questioned by R. Fagin and J. Halpern in 1987 when introducing their Logic for Local Reasoning [51]. In [51] it is argued that agents do not focus on all issues simultaneously and that saying that 'agent a believes p' rather means that in a certain perspective, when agent a considers the issues involving p, she believes p. All these different perspectives of an agent together then encompass what she knows and believes, instead of viewing

knowledge and beliefs of agents as absolute. This is based on the idea of agents as a *society-of-minds* [26, 43, 78]: they are a collection, *society*, of non-interacting belief clusters, *minds*. These clusters serve as frames or perspectives of the agent. Depending on the situation, the agent 'chooses' a cluster, reasons with what is available to her in this cluster and temporarily 'forgets' about the other clusters. Specific to the logic in [51] is that the different clusters are allowed to be contradictory with another, as long as they are consistent in themselves. Agents are therefore, within their frames, perfectly consistent reasoners, but they may be inconsistent globally.

A similar approach may benefit the logical model of ARG by letting the alignments used by agents correspond to beliefs in distinct, non-interacting clusters. Common to these clusters are then the ontologies of the agents. This approach to model agents' beliefs as distinct clusters could be a way to overcome the difference between adaptive agents and logical agents concerning local and global reasoning. Such a modification would let the logical agents use the alignments one by one instead of globally, just like adaptive agents.

4.2 Vocabulary Awareness

The second difference between adaptive agents and logical agents concerns the vocabularies used by agents in ARG to represent their knowledge, and more particularly the *awareness* agents have of these vocabularies. Adaptive agents use *different* and *dynamic* vocabularies to represent their knowledge and beliefs, while the logical agents *share* a *fixed* vocabulary.

In ARG, each agent has her own vocabulary to express her knowledge and beliefs. These vocabularies are typically not available to other agents because they are developed from different sources or from learning autonomously. Only the terms in another agent's vocabulary occurring in a correspondence of the alignment are accessible to the agents. That is, agents only know how to understand a class in other agents' ontologies if that class can be 'translated', via the alignment, to a class in their own ontology, and they are unaware of other classes. Thus, agents are aware of their own vocabulary and of the sub-vocabulary of other agents that is in correspondence with their own in the alignment.

Furthermore, vocabularies are dynamic for ARG. Agents may add classes to their ontologies when encountering new objects in the environment or may learn classes in other agents' ontologies when communicating with them. The latter occurs when agents repair their alignments and new correspondences are added that involve a class not in the alignment before. In both cases, this

means that agents may increase their awareness and therefore extend their vocabularies.

Logical agents, on the contrary, share a fixed vocabulary. This is because DEOL, but more generally DEL, is defined in function of a unique vocabulary, referred to as the set of propositions, or P, and it makes use of total valuation functions. These elements of DEL ensure that any communication or observation taking place, modeled by the dynamic modalities (announcements, radical and conservative upgrades), can be understood by all agents and agents can update or revise their knowledge and beliefs accordingly. However, this prevents agents to be aware of different vocabularies, or to extend them. Logical agents learn through re-organizing or eliminating possibilities. As a consequence, the logical agents have full awareness of *all* the propositions used currently, *and* in the future, by *all* agents. This means that any future evolution of the information the agents have, in the form of knowledge or belief or awareness, is restricted to the initial situation. Agents are therefore fully aware of the vocabularies used by other agents, even if it is not part of the alignment, and of the vocabulary that they will use in the future but do not yet use in the present moment. The vocabularies are fixed and shared among all agents.

Heterogeneous knowledge representations were assumed for ARG because in dynamic and open multi-agent systems, agents typically use different vocabularies to express their knowledge and beliefs and these vocabularies are required to continuously evolve [4, 46]. Alignments then define the part of the vocabularies the agents have access to (their *awareness*) and how to give meaning to it via semantic relations (their *knowledge* and *beliefs*). The distinction between awareness and knowledge or beliefs of agents, however, is not defined in DEL, nor DEOL. This causes any logical model based on DEL to be insufficient to model agents with heterogeneous knowledge representations based on different and dynamic vocabularies. In particular therefore, such a logical model is insufficient to model ARG. Hence, DEL has to be modified. For example, through dealing with awareness.

Previous work on awareness (see Section 2.4 on page 43) have concentrated on awareness of the truth value of a statement and not on awareness of the statement itself. This is because they are based on total valuation functions, on which awareness acts as a filter. This implies that, even though agents may use different vocabularies, raising the awareness of a proposition must come equipped with disclosing its underlying truth value, which is initiated by the valuation function. Hence, such a notion of awareness still prevents agents to openly evolve their vocabularies.

Therefore, to account for both agents with different vocabularies and dynamic vocabularies, awareness needs to be introduced differently. This

will be explored in Chapter 5.

4.3 Discarding Evidence

The third difference between adaptive agents and logical agents concerns the ability of agents to remember the evidence or focus on general conclusions. Adaptive agents use the objects but discard them after drawing general knowledge, while the logical agents cannot forget.

In ARG, agents use objects to test and evolve their alignments. They draw an object, communicate its class and in case of a failure, repair the alignment via adaptation operators. However, after applying an adaptation operator to the correspondence that caused the failure, they discard the object that led them to evolve [46, 49]. In other words, adaptive agents forget the individual cases because they focus on general conclusions. Even if memory has been added, the agents only record the faulty correspondence and not the class of the object drawn [49].

Logical agents, on the contrary, remember all the objects drawn and the classifications communicated. That is, they cannot forget and focus on more general conclusions like adaptive agents. This is because, when the object is drawn, its class is communicated, which is translated to the announcement $!C_b(o)$. This announcement causes the agents to acquire knowledge and DEL, and therefore DEOL, do not have a way to adjust and forget this knowledge.

Whether or not the agents remember all the individual cases or concentrate on general knowledge and beliefs plays a prominent role in the proof of incompleteness of the adaptation operators (Proposition 3.8 on page 70). This proof makes use of the fact that the logical agents acquire knowledge after the announcement $!C_b(o)$. More precisely, after applying $!C_b(o)$ to the translation $\tau(s)$ of an ARG state s and adaptation operator α, it holds that agent a knows to which class o belongs in agent b's ontology:

$$(\tau(s))^{\delta(\alpha)} \models K_a(C_b(o)) \tag{4.4}$$

On the contrary, this knowledge is not acquired by agent a in ARG because in ARG, agents evolve through adaptation operators and these adaptation operators only affect the alignments, i.e. beliefs, of agents. This means that if agent a did not know something before applying the adaptation operator, she also does not know it after applying it. In particular, this holds for $C_b(o)$. Therefore:

$$\tau(\alpha(s)) \not\models K_a(C_b(o)) \tag{4.5}$$

Equations 4.4 and 4.5 lead to the incompleteness of the adaptation operator (Proposition 3.8, page 70).

The fact that the adaptive agents do not remember the cases is not a problem for adaptive agents. The adaptive agents focus on the alignment and how to improve the alignment through communication, and only use the objects as tools for this purpose. In fact, the alignment alone is satisfactory for agents to communicate successfully: the experimental results in [46, 49] show that, through ARG, agents improve their alignments and achieve successful communication. Hence, they do not actually need to remember the individual objects that caused failures in the past to prevent the same failures from happing in the future. This is because the objects have led the agents to repair their alignments for exactly this purpose.

Even though the same failure is prevented, this does not mean the repair is successful immediately. For the adaptation operator `add` it may occur that the added correspondence will cause another failure in future rounds. More precisely, this happens when `add` is different from `addjoin`. In this situation, remembering the cases, the objects, will actually make the ontology and alignments locally inconsistent, leading to the logical translation of this state to be inconsistent and hence not have a model (see Proposition 3.2 on page 55). Therefore, in a way, not remembering the cases is an advantage for adaptive agents, as they are not 'lost' in this situation like logical agents but can continue to repair the alignment to reach successful communication.

The announcement $!C_b(o)$ is closely linked to incompleteness, but also to partial redundancy. The proof of partial redundancy is based on the fact that the logical agents may already deduce beliefs that correspond to the revised alignment without having to apply the adaptation operators. One may therefore think that the announcement, playing an important role in both the incompleteness and redundancy proofs, could have been avoided in the translation of ARG in the first place, but this is not the point because the current translation is indeed faithful to the communication taking place in ARG: agents announce to which class the object belongs, they use this fact to improve their alignment and then forget it. Instead of discarding the announcement, the missing ingredient in the translation is rather a form of forgetting.

The modification of the logic to account for forgetting is related to agent awareness. Once awareness is introduced and a raising awareness modality is defined, forgetting can be defined accordingly [37, 41]. Whether this is a reverse modality of raising awareness will be explored in Chapter 6.

4.4 Conclusion

In this chapter, the interpretation of the lack of formal properties as an indicator that the logical model is insufficient to model the knowledge evolution of agents in ARG is further explored. In particular, three differences between adaptive agents and logical agents have been identified that motivate different courses of action to modify the logic used to model ARG. These courses of action consist of (1) considering the knowledge and beliefs of agents not as absolute but as occurring in distinct clusters, allowing agents to reason locally over their alignments, (2) introducing a new notion of awareness for agents, enabling them to use different vocabularies that may openly evolve and (3) defining a forgetting modality to let agents focus on general knowledge. Through this, it is expected that the logical model will be closer resembling adaptive agents.

It must be noted that the differences presented in this chapter may not be a complete list of differences between adaptive and logical agents. They were found in an attempt to explain why the adaptive agents perform quite well according to experiments [46, 49] despite the lack of formal properties as proved in Chapter 3.

In the next chapter, vocabulary awareness is further explored in an attempt to finding a closer connection between experimental cultural knowledge evolution and a logical model of it. For simplicity, this is first explored for DEL, on which DEOL is based, but is then equivalently applied to DEOL.

Chapter 5

Agent Awareness

This chapter drops the assumption of full vocabulary awareness that is present in Dynamic Epistemic (Ontology) Logic (DE(O)L) in order to account for dynamic and open multi-agent systems that respect heterogeneity between agents, like the Alignment Repair Game (ARG). This is achieved through introducing a notion of awareness based on partial valuation functions and weakly reflexive relations. Partial valuation functions create a distinction between *uncertain* agents, agents that are aware of a proposition but do not know the truth value, and *unaware* agents, agents that do not consider the proposition at all. Weakly reflexive relations enable agents to use different vocabularies to represent their knowledge and beliefs. Together, they enable agents to use different vocabularies that can be extended when agents encounter new terms, either from the environment or from interacting with other agents.

In the following, first the use of partial valuation functions and weakly reflexive relations is motivated and how to model awareness accordingly is discussed. Properties of awareness are introduced, and knowledge and beliefs of agents are defined with respect to the awareness of agents. This leads to define Partial Dynamic Epistemic Logic (ParDEL). Dynamic modalities for raising public and private awareness are then introduced for ParDEL that duplicate the worlds in which the proposition awareness is raised of is undefined, and make it true in one while false in the other. As a consequence, raising awareness is disconnected from learning truth: after awareness is raised, unaware agents become uncertain. Finally, it is discussed how the notion of awareness could benefit DEL more broadly by enabling private upgrades to take place.

5.1 Why DEL is insufficient to model dynamic and open multi-agent systems

Dynamic Epistemic Logic (DEL) is a rich framework for analysing epistemic and doxastic changes under dynamic upgrades. It has been widely used as a formal framework to model agent communication [14, 42, 81], belief revision [12] and agent interaction [13]. However, there is a category of multi-agent systems left unaddressed: *dynamic* and *open* multi-agent systems, respecting heterogeneity between agents. The Alignment Repair Game (ARG) is such a system.

There are two problems when it comes to using DEL to model dynamic and open multi-agent systems: (1) agents cannot use their *own* vocabularies in DEL, and (2) agents cannot *extend* nor *shrink* their vocabularies. There are three aspects of DEL that cause these problems: (a) the set of propositions P is fixed and not open to evolve, (b) valuation functions are total functions and (c) accessibility relations are reflexive. Here, we discuss the alternatives: a dynamic set of propositions P enables open evolution, partial valuation functions let agents be aware of a subset of P, and weakly reflexive relations allow agents to have different awareness.

5.1.1 A Dynamic Set of Propositions

Using a fixed set of propositions P enforces the vocabularies of agents to be restricted to it. Even considering the other two issues aside, this means that agents could evolve their vocabularies only within P and not openly. Furthermore, once their vocabularies contains all of P, they can no longer acquire new terms.

In an analogy to human language, where P would be the set of all the words in every language available, this means that humans could extend their vocabularies with words in another language or that a meaning of a term could change, but no new words can be introduced. Hence, the languages themselves cannot evolve. Both for humans and agents, this is not desirable nor realistic. Therefore the set of propositions P needs to be able to evolve, requiring a dynamic approach.

5.1.2 Partial Valuation Functions

Solely letting P be a dynamic set of propositions, however, does not solve the problem that DEL is not sufficient to model dynamic and open multi-agent systems where agents use heterogeneous sources. It still causes agents to use

the same vocabulary, namely P, and to evolve it simultaneously.

The reason is that valuation functions are total. Total valuation functions cause the vocabularies of agents to be equivalent to P. This is because it enforces each proposition to have a valuation at each world, therefore enforcing agents to be aware of all of them. With partial valuation functions, on the contrary, propositions may either be true, false, or *undefined*. The latter occurs when the propositions are not interpreted by the valuation function at a certain world – they do not belong to the domain $Dom(V_w)$ of the valuation function V at world w. When each world has a different partial valuation function specifying which propositions are evaluated (they belong to the domain) and how they are evaluated (whether they are assigned truth or falsity), agents can be aware of a subset of P.

5.1.3 Weakly Reflexive Relations

It is still not enough to let P be a dynamic set of propositions and valuation functions V to be partial functions. When the accessibility relations are reflexive, agents still share the same vocabulary. Reflexive relations cause agents to have access to the same worlds, and therefore be aware of the same vocabularies, defined as the domain at these worlds. Therefore we drop reflexivity as an assumption on the model structures. This allows for situations in which two agents a and b can have access to different worlds v and u, respectively, from w in a pointed model $\langle \mathcal{M}, w \rangle$ and $Dom(V_v) \neq Dom(V_u)$, i.e. the agents can use different vocabularies.

We do not drop reflexivity completely, just globally. For each agent, reflexivity is still needed at some worlds to determine her awareness, where the awareness of an agent is defined as the set of propositions defined in all the worlds accessible by that agent with reflexivity for her. Furthermore, some form of reflexivity is necessary to maintain the usual properties of knowledge and belief, and in particular factivity of knowledge. We call this other form of reflexivity *weak reflexivity*. In short, weak reflexivity enforces that whenever agents have a reflexive accessibility relation at w and they can access another world v, they also have a reflexive accessibility relation at v.

5.2 A Definition of Awareness

The accessibility relations therefore indicate which worlds the agents are aware of – those they can access (via R or R^{-1}) where reflexivity is satisfied – and the partial valuation functions determine which propositions the agents are aware of – those propositions belonging to the domain of these

worlds. Both R and R^{-1} are taken into account so to make R act as a plausibility relation, comparing the plausibility of two worlds, but now restricted to reflexivity.

Definition 5.1 (Agent Awareness)**.** Let W be a non-empty set of worlds, P be a countable, non-empty set of propositions, $\mathcal{A}$ be a finite, non-empty set of agents, and $\{V_w\}_{w \in W}$ be a set of partial valuation functions, one for each world. Then, if $R_a \subseteq W \times W$ is an accessibility relation for agent a, we say that a is *aware* at w if and only if wR_aw and the *awareness* (or vocabulary) of agent a at w is defined as:

$$AW_a(w) = \bigcup_{\{v \in W \ \mid \ w(R_a \cup R_a^{-1})v \wedge vR_av\}} Dom(V_v) \tag{5.1}$$

Whenever w is clear from the context as the pointed world in a pointed model $\langle \mathcal{M}, w \rangle$, we will also use AW_a for $AW_a(w)$ to denote the awareness of agent a in the pointed model.

This definition enables agents to use their own signatures to represent their knowledge and beliefs and to adapt or extend it through learning, making it possible for agents to openly evolve their knowledge and beliefs without fixing the evolution in the initial setting. For example, in Figure 5.1, the awareness of agent a is $\{p\}$ while the awareness of agent b is $\{q\}$.

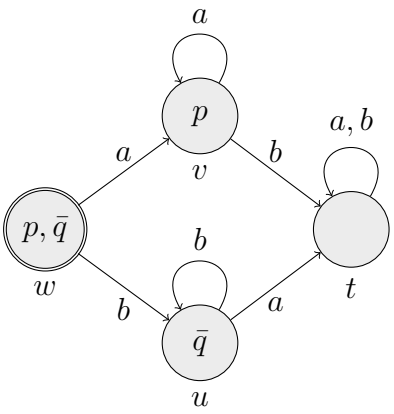

Figure 5.1: Two agents, a and b, with different, disjoint, awareness $\{p\}$ and $\{q\}$, respectively.

The consequence of considering partial valuation functions is that lack of truth and falsity do no longer coincide. Instead, there is a third option when propositions are *undefined*. Whenever this happens for a proposition

p at a world accessible by an agent, this agent is said to be *unaware* of p. Unawareness is different from uncertainty, also called ignorance. The latter occurs when agents have no information about the truth value of a proposition, they do not know nor believe the proposition, nor the truth value. Whereas unawareness occurs when agents do not consider a proposition at all. This means that p is undefined in the worlds accessible by the agent. On the contrary, uncertain agents have access to at least one world in which p is true and at least one world in which p is false, while they are both considered equally plausible, see Figure 5.2.

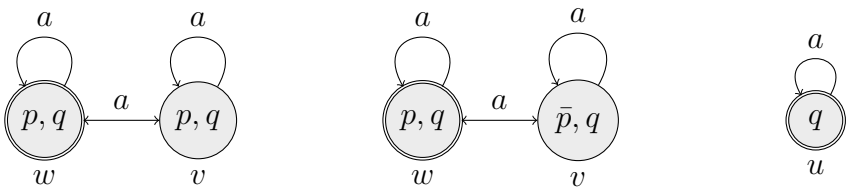

Figure 5.2: Agent a is certain about p ($\models K_a p$, left), uncertain about p ($\models K_a(p \vee \neg p)$, middle) and unaware of p ($\not\models K_a p$ and $\not\models K_a p$, right). In all cases, agent a knows q.

When unawareness of a proposition means that the agents do not consider a proposition, awareness means that agents do consider the proposition, and hence also a truth value of it. This means that awareness of a proposition implies that agents are at least uncertain about the proposition.

By allowing propositions to be true, false and undefined, agents can use their own vocabularies to express their knowledge and beliefs. A vocabulary for an agent a is a set of the propositions defined in all the worlds accessible by a where reflexivity is satisfied. 'Using their own vocabulary' here means that agents are aware of a different vocabulary, where the awareness of an agent is determined by the propositions defined in the worlds accessible that satisfy reflexivity.

5.3 Properties of Awareness

We require that the awareness of agents is constant over their accessibility. This is to ensure that agents do not consider it possible that their awareness is greater or smaller than it actually is. In other words, their awareness is constant over their considerations. Similar properties for different notions of awareness were already motivated in [51, 56]: in [51], awareness is assumed

to not be able to decrease under evolution (i.e. there is no "forgetting") and in [56] awareness is considered constant for all the worlds the agent has access to. Compared to their work, the properties of awareness we introduce here are not different, but the framework for awareness itself (via partial valuation functions and weakly reflexive relations).

Requiring that awareness is constant over accessibility does not imply that it is fixed: agents can extend their awareness through model-changing dynamic upgrades that is not restricted to the set of propositions P, which we will define Section 5.6.

Letting agent awareness be constant over their accessibility comes two-fold:

- whenever there is a reflexive relation for an agent a from a world w to w, then for any v that is also accessible from w for a, there is a reflexive relation from v to v (*weak reflexivity*), and
- the domains of two worlds v, u that are both accessible by the same agent from a single world w are equal (*consideration consistency*).

Weak reflexivity ensures that if an agent is aware at a world w, she will remain aware at all the worlds she can reach from w, see Figure 5.3.

Definition 5.2 (Weak Reflexivity)**.** Let $\mathcal{A}$ be a finite, non-empty set of agents, W a non-empty set of worlds and let $a \in \mathcal{A}$ with accessibility relation $R_a \subseteq W \times W$. Then R_a is *weakly reflexive* if $\forall w, v \in W$:

$$wR_aw \wedge wR_av \Rightarrow vR_av \tag{5.2}$$

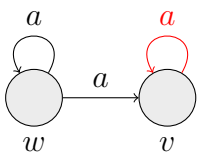

Figure 5.3: A visualization of weak reflexivity: if the black arrows hold for a certain agent a, then there must also be the red arrow for a.

Consideration consistency ensures that the awareness of agents, i.e. the propositions that are defined in a world with reflexivity, is constant over accessibility, see Figure 5.4.

Definition 5.3 (Consideration Consistency)**.** Let W be a non-empty set of worlds, P be a countable, non-empty set of propositions, $\mathcal{A}$ be a finite, non-empty set of agents, $\{V_w\}_{w \in W}$ be a set of partial valuation functions, and let $a \in \mathcal{A}$ with accessibility relation $R_a \subseteq W \times W$. Then $\{V_w\}_{w \in W}$ satisfies *consideration consistency* if $\forall w, v, u \in W$:

$$wR_av \wedge wR_au \Rightarrow Dom(V_v) = Dom(V_u) \tag{5.3}$$

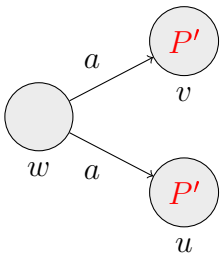

Figure 5.4: A visualization of consideration consistency: if the black arrows hold for a certain agent a, and the domains are $Dom(V_v)$ and $Dom(V_u)$, then these domains are equivalent, here drawn to be P'.

Consideration consistency does not only apply to worlds w, v, u such that wR_av and wR_au (hence $Dom(V_v) = Dom(V_u)$), but also stipulates that if wR_aw and wR_av then $Dom(V_w) = Dom(V_v)$. Therefore, when combined with weak reflexivity, it enforces that agents are consistent in their considerations: if an agent a considers a proposition p (or its negation) to be true at a world w (i.e. $p \in Dom(V_w)$ and wR_aw), she considers p to be true or false at every world v she can reach via R_a from w (i.e. also $p \in Dom(V_v)$ and vR_av), and vice-versa. This is independent from the truth value of p – it only requires that p is *assigned* a truth value.

Proposition 5.1 (Awareness constant over accessibility)**.** Let W be a non-empty set of worlds, P be a countable, non-empty set of propositions, $\mathcal{A}$ be a finite, non-empty set of agents, and let $\{V_w\}_{w \in W}$ be a set of partial valuation functions satisfying consideration consistency. If $a \in \mathcal{A}$ with accessibility relation $R_a \subseteq W \times W$ that is weakly reflexive, then it holds that $\forall w \in W$: if wR_aw then $\forall v \in W$ such that wR_a^*v: vR_av and $AW_a(v) = AW_a(w)$, where R_a^* is the transitive closure of R_a.

Proof. Follows directly by Definition 5.2 and 5.3 for weak reflexivity and consideration consistency. □

Lastly, we require that agents cannot reason about the knowledge or beliefs of other agents when it involves a proposition they are not aware of themselves. This is called specification:

- the domain of the valuation function cannot increase over accessibility (*specification*).

Specification avoids the situation in which, for example, $K_a K_b p$ but $p \notin AW_a$. Note that when wR_aw and wR_av, by consideration consistency, $Dom(V_v) = Dom(V_w)$ and therefore trivially $Dom(V_v) \subseteq Dom(V_w)$. See Figure 5.5 for a visualization of the property.

Definition 5.4 (Specification)**.** Let W be a non-empty set of worlds, P be a countable, non-empty set of propositions, $\{V_w\}_{w \in W}$ be a set of partial valuation functions, and let $a \in \mathcal{A}$ with accessibility relation $R_a \subseteq W \times W$. Then $\{V_w\}_{w \in W}$ satisfies *specification* if $\forall w, v \in W$:

$$wR_av \Rightarrow Dom(V_v) \subseteq Dom(V_w) \tag{5.4}$$

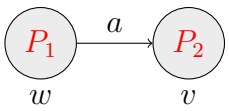

Figure 5.5: A visualization of specification: if the black arrow from w to v holds for a certain agent a, and the domains are $Dom(V_w) = P_1$ and $Dom(V_v) = P_2$, then these $P_2 \subseteq P_1$.

It follows that if two agents both have reflexivity satisfied at w, their awareness is the same, i.e. $AW_a(w) = AW_b(w)$.

The three components, weak reflexivity, consideration consistency and specification, ensure that agents are consistent in their considerations, while enabling agents to be aware of different vocabularies.

5.4 Knowledge and Belief

Knowledge and beliefs of agents are defined with respect to the accessibility relations as usual. However, with reflexivity not satisfied globally, it comes with the additional requirement that agents need to be aware at the worlds reached via accessibility. Knowledge is defined as everything that is true in all accessible worlds for an agent in which reflexivity is satisfied for that

agent, and belief is defined as everything that is true in the maximal worlds with respect to the accessibility relation for an agent in which reflexivity is satisfied for that agent. These notions can also be captured by *epistemic* and *doxastic* relations. They are defined with respect to the *aware cell* of an agent. The aware cell of an agent at a world are all the worlds accessible by the agent that satisfy reflexivity for her, see Figure 5.6 for an example.

Definition 5.5 (Aware Cell)**.** Let $\mathcal{A}$ be a finite, non-empty set of agents, W be a non-empty set of worlds and $a \in \mathcal{A}$ be an agent with accessibility relation $R_a \subseteq W \times W$ that is well-founded, locally connected, weakly reflexive and transitive. The *aware cell* of agent a at world $w \in W$, denoted by $||w||_a$, is the set of worlds that are accessible via the transitive closure of R_a and R_a^{-1} and in which reflexivity is satisfied.

$$||w||_a = \{v \in W \mid w(R_a \cup R_a^{-1})^* v \text{ and } vR_av\} \tag{5.5}$$

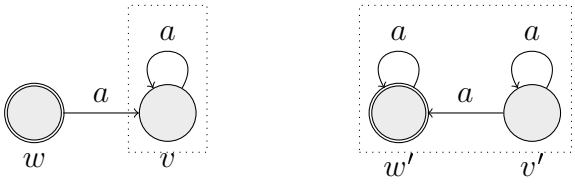

Figure 5.6: The aware cells are as follows: $||w||_a = ||v||_a = \{v\}$ and $||w'||_a = ||v'||_a = \{w', v'\}$ (marked by dotted boxes).

Awareness of an agent a at world w, $AW_a(w)$ (Definition 5.1 on page 88), can now also be defined with respect to the aware cell of agents.

Proposition 5.2. Given a finite, non-empty set $\mathcal{A}$ of agents and a non-empty set W of worlds, then for all $w \in W$ and all $a \in \mathcal{A}$: if $R_a \subseteq W \times W$ is an accessibility relation that is well-founded, locally connected, weakly reflexive and transitive, then

$$AW_a(w) = \bigcup_{v \in ||w||_a} Dom(V_v) \tag{5.6}$$

Proof. Let w be a world and $a \in \mathcal{A}$ an agent. Then the awareness of a is defined as $AW_a(w) = \bigcup_{\{v \mid w(R_a \cup R_a^{-1})v \wedge vR_av\}} Dom(V_v)$ (Definition 5.1 on page 88). But since transitivity is satisfied, the worlds v such that $w(R_a \cup R_a^{-1})v$ are exactly those worlds such that $w(R_a \cup R_a^{-1})^* v$. Hence $v \in ||w||_a$ and therefore $AW_a(w) = \bigcup_{v \in ||w||_a} Dom(V_v)$. □

Given a set of partial valuation functions that satisfy consideration consistency and a relation R_a that is well-founded, locally connected, weakly reflexive and transitive, it follows that within an aware cell, the awareness of agents is constant.

Proposition 5.3 (Awareness constant in aware cells)**.** Given a finite, non-empty set $\mathcal{A}$ of agents and a non-empty set W of worlds, then for all $w \in W$ and all $a \in \mathcal{A}$: if V is a partial valuation function that satisfies consideration consistency and $R_a \subseteq W \times W$ is an accessibility relation that is well-founded, locally connected, weakly reflexive and transitive, then $\forall u, v \in ||w||_a$: $AW_a(v) = AW_a(u)$.

Proof. Follows directly from Proposition 5.1 (page 91). □

We also use $Dom(V_w)$ when we talk about the awareness of agent a at world w.

Epistemic and doxastic relations are defined with respect to the aware cell of an agent.

Definition 5.6 (Epistemic and Doxastic Relations)**.** Let $\mathcal{A}$ be a finite, non-empty set of agents, W be a non-empty set of worlds and let $a \in \mathcal{A}$ be an agent with accessibility relation $R_a \subseteq W \times W$ that is well-founded, locally connected, weakly reflexive and transitive. Then the epistemic ($\sim_a$) and doxastic ($\rightarrow_a$) relations are defined as follows:

$$w \sim_a v \text{ iff } v \in ||w||_a \tag{5.7}$$

$$w \rightarrow_a v \text{ iff } v \in Max_{R_a}||w||_a \tag{5.8}$$

Since for constructing $\sim_a$ and $\rightarrow_a$, we require R_a to be well-founded, locally connected, weakly reflexive and transitive, it follows that, within aware cells of an agent a, $\sim_a$ and $\rightarrow_a$ satisfy the usual properties for epistemic and doxastic relations. in

Proposition 5.4. Let $\mathcal{A}$ be a finite, non-empty set of agents, W be a non-empty set of worlds and let $a \in \mathcal{A}$ be an agent with accessibility relation $R_a \subseteq W \times W$ that is well-founded, locally connected, weakly reflexive and transitive. Then within $||w||_a$, $\sim_a$ is reflexive, transitive and symmetric and $\rightarrow_a$ is transitive, serial and Euclidean.

Proof. The only difference with the plausibility relation $\leqslant_a$ for DEL (Definition 2.21 on page 31) is that R_a is not reflexive, but weakly reflexive. Then, by Definition 5.5 (page 93) of aware cells, $\forall w$ and $\forall v \in ||w||_a$: vR_av. Hence, within $||w||_a$, the relation R_a is a well-founded locally connected pre-order

and it therefore follows that $\sim_a$ and $\rightarrow_a$ satisfy the same properties as the epistemic and doxastic relations for DEL [42]: $\sim_a$ is reflexive, transitive and symmetric and $\rightarrow_a$ is transitive, serial and Euclidean. □

Therefore, within the aware cells of agents, knowledge and belief satisfy the usual axioms as for DEL: $S5$ and $KD45$, respectively. However, outside the aware cells, it is not given that knowledge is factive. This means that the T axiom is no longer valid, but is only valid within the aware cells of agents. This behavior is due to the weak reflexivity condition as defined in Definition 13 (page 14). As a consequence, the knowledge operator is an intermediate operator between $S5$ and $KD45$: K_a satisfies $S5$ within the aware cell of agent a, but only satisfies $KD45$ outside the aware cell.

In practice, such a knowledge operator may be thought of as "subjective knowledge": it takes the perspective of the agent in question and what is known by her.

5.5 Partial Dynamic Epistemic Logic

Here, we introduce Partial Dynamic Epistemic Logic (ParDEL) as an extension of DEL with weakly reflexive relations and partial valuations satisfying consideration consistency and specification. The syntax of ParDEL is equivalent to that of DEL (Definition 2.14 on page 26).

Definition 5.7 (Syntax of ParDEL)**.** Given a countable, non-empty set P of propositions and a finite, non-empty set $\mathcal{A}$ of agents, the *syntax*, $\mathcal{L}_{ParDEL}$, of (multi-agent) Partial Dynamic Epistemic Logic (ParDEL) is defined in the following way.

$$\phi ::= p \mid \phi \wedge \psi \mid \neg\phi \mid K_a\phi \mid B_a\phi \mid [\dagger\phi]\psi$$

where $p \in P$ is a proposition, K_a and B_a are the knowledge and belief operators for each agent a, and $\dagger\phi$ with $\dagger \in \{!, \Uparrow, \uparrow\}$ the dynamic upgrades (announcements, radical and conversative upgrades).

ParDEL frames are DEL frames but satisfy weak reflexivity instead of reflexivity in order to allow agents to use different signatures to represent their knowledge and beliefs.

Definition 5.8 (ParDEL Frames)**.** Given a finite, non-empty set $\mathcal{A}$ of agents, *frame* of (multi-agent) ParDEL is a pair $\mathfrak{F} = \langle W, (R_a)_{a\in\mathcal{A}}\rangle$ where

- W is a non-empty set of worlds, and

- $(R_a)_{a \in \mathcal{A}} : \mathcal{A} \to \mathcal{P}(W \times W)$ are the accessibility relations on W, one for each agent, that are well-founded, locally connected, weakly reflexive and transitive.

ParDEL models are ParDEL frames equipped with a partial valuation function satisfying consideration consistency and specification.

Definition 5.9 (ParDEL Models)**.** Given a countable, non-empty set P of propositions and a finite, non-empty set $\mathcal{A}$ of agents, a *model* of (multi-agent) ParDEL is a pair $\mathcal{M} = \langle \mathfrak{F}, V \rangle$ where

- $\mathfrak{F} = \langle W, (R_a)_{a \in \mathcal{A}} \rangle$ is a ParDEL frame, and
- $V : W \to (P \to \{0, 1\})$ is a *partial valuation function* that assigns to each world $w \in W$ a partial function $V_w : P \to \{0, 1\}$ satisfying consideration consistency and specification.

A *pointed ParDEL model* is a pair $\langle \mathcal{M}, w \rangle$ where $\mathcal{M}$ is a ParDEL model and $w \in W$.

Satisfiability for ParDEL is considered with respect to a pointed model $\langle \mathcal{M}, w \rangle$ which associates a ParDEL model $\mathcal{M}$ with a world $w \in W$. Since lack of truth and falsity do not coincide with partial valuations, two relations are specified: one for *verification* ($\models$) and one for *falsification* ($\dashv$).

Different from other approaches to defining partial valuations functions for logic [64], is that we define falsification of a conjunction whenever both conjuncts are defined and at least one of them is falsified. This is required to prevent agents from gaining knowledge or belief that a conjunction is false whenever they know that one of the conjuncts is false, but are unaware of the other conjunct. In other words, it enforces that the knowledge and beliefs of agents are restricted to their awareness.

Definition 5.10 (Satisfiability for ParDEL)**.** Satisfiability for ParDEL extends that of DEL (Definition 2.18 on page 29) and is defined as:

$$
\begin{array}{ll}
\mathcal{M}, w \models p & \text{iff } V_w(p) = 1 \\
\mathcal{M}, w \models \phi \wedge \psi & \text{iff } \mathcal{M}, w \models \phi \text{ and } \mathcal{M}, w \models \psi \\
\mathcal{M}, w \models \neg\phi & \text{iff } \mathcal{M}, w \dashv \phi \\
\mathcal{M}, w \models K_a\phi & \text{iff } \forall v \text{ s.t. } w \sim_a v : \mathcal{M}, v \models \phi \\
\mathcal{M}, w \models B_a\phi & \text{iff } \forall v \text{ s.t. } w \to_a v : \mathcal{M}, v \models \phi \\
\mathcal{M}, w \models [\dagger\phi]\psi & \text{iff } \mathcal{M}^{\dagger\phi}, w \models \psi
\end{array}
$$

for verification ($\models$) and:

$$
\begin{array}{ll}
\mathcal{M}, w \dashv p & \text{iff } V_w(p) = 0 \\
\mathcal{M}, w \dashv \phi \wedge \psi & \text{iff } \mathcal{M}, w \dashv \phi \text{ and } \mathcal{M}, w \vDash \psi, \\
 & \quad \text{or } \mathcal{M}, w \vDash \phi \text{ and } \mathcal{M}, w \dashv \psi, \\
 & \quad \text{or } \mathcal{M}, w \dashv \phi \text{ and } \mathcal{M}, w \dashv \psi \\
\mathcal{M}, w \dashv \neg\phi & \text{iff } \mathcal{M}, w \vDash \phi \\
\mathcal{M}, w \dashv K_a\phi & \text{iff } \exists v \text{ s.t. } w \sim_a v : \mathcal{M}, v \dashv \phi \\
\mathcal{M}, w \dashv B_a\phi & \text{iff } \exists v \text{ s.t. } w \rightarrow_a v : \mathcal{M}, v \dashv \phi \\
\mathcal{M}, w \dashv [\dagger\phi]\psi & \text{iff } \mathcal{M}^{\dagger\phi}, w \dashv \psi
\end{array}
$$

for falsification ($\dashv$), with $\dagger \in \{!, \Uparrow, \uparrow\}$.

Again, a set of formulas is said inconsistent if there does not exist a pointed model verifying it. In the following, we say that a formula ϕ is a consequence of a set of formulas Γ (written $\Gamma \vDash \phi$) if every pointed model $\langle \mathcal{M}, w \rangle$ verifying all formulas of Γ, also verifies ϕ.

Whenever a proposition p does not belong to the domain of the valuation function at a world w, i.e. $p \notin Dom(V_w)$, it holds that $\mathcal{M}, w \nvDash p$ and $\mathcal{M}, w \ndashv p$. Hence, p is neither true nor false.

As disjunctions and implications can be defined using conjunctions and negations, the falsification clause for conjunctions also affects their satisfiability.

Example 5.1. Since $\phi \vee \psi$ is the abbreviation for $\neg(\neg\phi \wedge \neg\psi)$:

$$
\begin{array}{ll}
\mathcal{M}, w \vDash \phi \vee \psi & \text{iff } \mathcal{M}, w \vDash \phi \text{ and } \mathcal{M}, w \dashv \psi, \\
 & \quad \text{or } \mathcal{M}, w \dashv \phi \text{ and } \mathcal{M}, w \vDash \psi, \\
 & \quad \text{or } \mathcal{M}, w \vDash \phi \text{ and } \mathcal{M}, w \vDash \psi \\
\mathcal{M}, w \dashv \phi \vee \psi & \text{iff } \mathcal{M}, w \dashv \phi \text{ and } \mathcal{M}, w \dashv \psi
\end{array}
$$

Of course, the different clause for the falsification of the conjunction also means that some validities are lost compared to [64]. For example, $p \vee \neg p$ is no longer valid, because it is not satisfied at pointed models $\langle \mathcal{M}, w \rangle$ where $p \notin Dom(V_w)$. This is justified because we do not want agents to know or believe $p \vee \neg p$ if they are not aware of p.

There is a link between intuitionism [27] and awareness, in particular the way satisfiability is defined for ParDEL. Example 5.1 shows that the truth of disjunctions is defined constructively: $\phi \vee \psi$ is true as long as *both* disjoints are defined, of which at least one is true. In particular, satisfiability for

ParDEL follows that of Weak Kleene Logic [67]. In this sense, we may say that awareness is constructive. Other than that, however, are awareness and intuitionism rather orthogonal: intuitionism drops the law of the excluded middle, dissociating ϕ and $\neg\phi$, whereas awareness dissociates being aware of ϕ and knowing the truth value of ϕ and $\neg\phi$.

5.6 Raising Awareness

Partial valuation functions allow agents to use different signatures to represent their knowledge. We have seen how this can be used to describe awareness and unawareness of agents when we replace reflexivity by weak reflexivity and which properties of awareness are natural to require: consideration consistency and specification. Here, we discuss the dynamics of raising awareness. Because even though awareness of agents is constant over their accessibility, partial valuations enable to define a model-changing dynamic upgrade allowing agents to extend their awareness.

We extend the syntax of ParDEL with a dynamic modality $+p$ for raising the awareness of propositions p. We call the logic obtained *Partial Dynamic Epistemic Logic with raising awareness* (ParDEL+).

Definition 5.11 (Syntax of ParDEL+)**.** Given a countable, non-empty set P of propositions and a finite, non-empty set $\mathcal{A}$ of agents, the *syntax*, $\mathcal{L}_{ParDEL+}$, of (multi-agent) Partial Dynamic Epistemic Logic with raising awareness (ParDEL+) is defined in the following way.

$$\phi ::= p \mid \phi \wedge \psi \mid \neg\phi \mid K_a\phi \mid B_a\phi \mid [\dagger\phi]\psi \mid [+p]\phi$$

where $p \in P$ is a proposition, K_a and B_a are the knowledge and belief operators for each agent a, $\dagger\phi$ with $\dagger \in \{!, \Uparrow, \uparrow\}$ the dynamic upgrades announcements, radical and conversative upgrades, and $+p$ is a raising awareness operation.

Frames and models of ParDEL+ are equivalent to frames and models of ParDEL (Definitions 5.8 and 5.9 on page 95). Again, pointed models of ParDEL+ are pairs $\langle \mathcal{M}, w \rangle$ where $\mathcal{M}$ is a ParDEL+ model and w is a world in $\mathcal{M}$. In the following, we also use ParDEL(+) to denote models of both ParDEL and ParDEL+.

When we introduced awareness by means of partial valuation functions and weakly reflexive relations, we have discussed how to interpret awareness of a proposition: if an agent is aware of a proposition, she considers a truth value of the proposition. This means that awareness at least implies ignorance (but agents may additionally have knowledge or beliefs). Therefore,

becoming ignorant is the minimal way to raise awareness without disclosing truth.

To raise the awareness of a proposition p, the valuation function of each world in which p does not belong to the domain are extended to define p. In order to solely raise awareness, i.e. to perform it without disclosing truth values, each of these worlds is duplicated, accessibility to and from duplicated worlds being preserved, and p is made true in one world and false in the other. Other than the different valuation of p (and related sentences), the two duplicated worlds are indifferent: satisfying the same valuations and relations. This ensures that agents become aware of the proposition without learning the truth value. After raising the awareness of p, agents consider equally plausible that p is true or that p is false, see Figure 5.7. Therefore, raising the awareness of p transforms unaware agents into uncertain agents: agents who are aware of p but do not know whether p is true or false.

When raising the awareness of p, we categorize the worlds of the model into three sets: worlds in which p is true ($W|_p = \{w \in W \mid V_w(p) = 1\}$), worlds in which p is false ($W|_{\neg p} = \{w \in W \mid V_w(p) = 0\}$) and worlds in which p is undefined ($W \backslash (W|_p \cup W|_{\neg p})$). In this way, the worlds in which p is undefined can be identified and duplicated.

Definition 5.12 (Raising Awareness ($+p$))**.** Let $\mathcal{M} = \langle W, (R_a)_{a \in \mathcal{A}}, V \rangle$ be a ParDEL+ model and let $p \in P$ be a proposition. Then $+p$ is a model transformer $+p : \mathcal{M} \mapsto \mathcal{M}^{+p}$ where $\mathcal{M}^{+p}$ is the triple $\langle W^{+p}, (R_a^{+p})_{a \in \mathcal{A}}, V^{+p} \rangle$ defined by:

- $W^{+p} = W|_p \times \{1\} \ \cup \ W|_{\neg p} \times \{0\} \ \cup \ W \backslash (W|_p \cup W|_{\neg p}) \times \{0, 1\}$
- $\langle w, i \rangle R_a^{+p} \langle v, j \rangle$ iff $w R_a v$
- $V^{+p}_{\langle w,i \rangle}(q) = \begin{cases} V_w(q) & \text{if } q \neq p \\ i & \text{otherwise} \end{cases}$

The new valuation function corresponds to the old one in the case that p was defined: $V^{+p}_{\langle w,i \rangle}(p) = i = V_w(p)$ because only in this case $\langle w, 1 \rangle \in W^{+p}$, where $i = 1$ if $V_w(p) = 1$ and $i = 0$ if $V_w(p) = 0$. For worlds where p is undefined, i.e. $p \notin Dom(V_w)$, raising awareness of p maps w to both $\langle w, 1 \rangle$ and $\langle w, 0 \rangle$, in which p is made true and false, respectively.

Because worlds may be duplicated, satisfiability for $[+p]\phi$ is defined for all $\langle w, i \rangle \in W^{+p}$ with $i \in \{0, 1\}$ for verification, and for some $\langle w, i \rangle \in W^{+p}$ with $i \in \{0, 1\}$ for falsification.

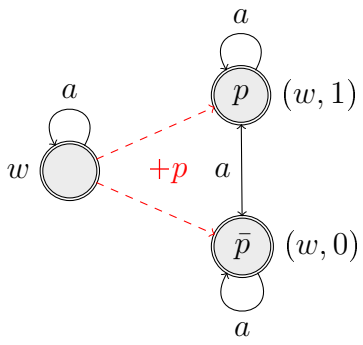

Figure 5.7: Raising the awareness of p, $+p$. The red dashed lines indicate how the world w from the model on the left is mapped to both $(w, 1)$ and $(w, 0)$ in the model on the right. p ($\bar{p}$) written inside a world w means that $V_w(p) = 1$ ($V_w(p) = 0$), and whenever neither p nor $\bar{p}$ is written, it means that p does not belong to $Dom(V_w)$.

Definition 5.13 (Satisfiability for ParDEL+)**.** Satisfiability for ParDEL+ extends that of ParDEL (Definition 5.10 on page 96) as follows:

$$\mathcal{M}, w \models [+p]\phi \quad \text{iff} \quad \forall \langle w, i\rangle \in W^{+p} \text{ s.t. } i \in \{0, 1\} : \mathcal{M}^{+p}, \langle w, i\rangle \models \phi$$
$$\mathcal{M}, w \dashv [+p]\phi \quad \text{iff} \quad \exists \langle w, i\rangle \in W^{+p} \text{ s.t. } i \in \{0, 1\} : \mathcal{M}^{+p}, \langle w, i\rangle \dashv \phi$$

See Figure 5.8 for an example of raising awareness of a proposition p when one agent is aware of p, but the other agent is not. Naturally, whenever p already belonged to the domain of the valuation function for a world w, raising the awareness of p does not affect w.

In Figure 5.8, we can also see that knowledge is no longer factive, *except* within the aware cells of agents: it holds that $(w, 0) \models K_a p$ implies $(w, 0) \models p$ because $(w, 0) \in ||(w, 0)||_a$, hence K_a is factive at $(w, 0)$. However, $(w, 0) \models K_b \neg K_a p$ holds, but $(w, 0) \not\models \neg K_a p$ because $(w, 0) \notin ||(w, 0)||_b$. Therefore K_b is not factive at $(w, 0)$. Hence, the knowledge operator considered is an operator acting between $S5$ and $KD45$, and we may call this "subjective knowledge". From the perspective of agent b, it is not the case that agent a knows p because b does not know that agent a was aware of p before the public raising awareness took place, whereas from agent a knows p.

Raising awareness is not restricted to the set of propositions P. We can use the definition to raise the awareness of $q \notin P$ and extend P with q. In this case, q will be undefined at every world of the model, leaving each world to be duplicated by $+q$. This means that P is not fixed in the initial setting.

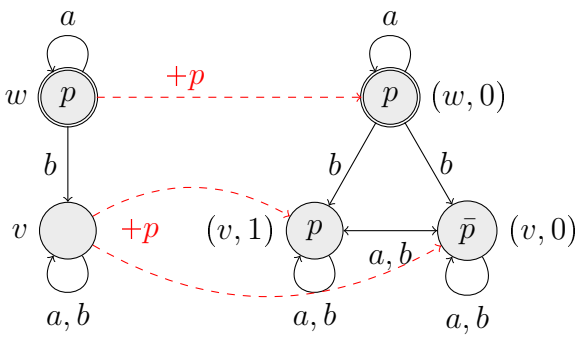

Figure 5.8: Raising the awareness of p, $+p$. The red dashed lines indicate how the world w from the model on the left is mapped to $(w, 1)$ in the model on the right and v to both $(v, 0)$ and $(v, 1)$.

5.7 Announcements, Radical and Conservative Upgrades on ParDEL+

With a notion of awareness and the possibility for propositions to be true, false and undefined, we have to discuss how the typical dynamic upgrades of DEL act on ParDEL+: announcements, radical and conservative upgrades.

There are three ways: (1) either it is required that agents are aware of the proposition that occurs in the upgrade, or (2) the upgrade only affects the worlds in which the proposition is defined and preserves those where it is undefined, or (3) the upgrade raises awareness first. For announcements, the first would require $!p$ to only be applied to models $\mathcal{M}$ such that $\mathcal{M}^{+p} = \mathcal{M}$. The second does not require awareness of all the agents, but deletes $\neg p$ worlds and not worlds in which p is undefined. This means that agents who are aware of p, learn the truth value, but other agents, who are unaware of p, cannot 'understand' the announcement and discard it. Finally, the third raises awareness while announcing.

Because the second case is most general (still, awareness *could* be raised before the upgrade), this is adopted for ParDEL+. With regards to the model transformers (Definition 2.22 on page 32), what changes are the definitions of radical and conservative upgrades: they change the ordering between worlds whenever they both verify or falsify ϕ. This ensures that the properties of ParDEL+ are preserved, in particular specification.

Announcement ($!\phi$) deletes all '$\neg\phi$'-worlds from the model, i.e. $W^{!\phi} = W \backslash \{w \mid \mathcal{M}, w \nVdash \phi\}$, $w \leqslant_a^{!\phi} v$ iff $w \leqslant_a v$ and $w, v \in W^{!\phi}$, $V^{!\phi}(p) = V(p) \cap W^{!\phi}$;

Radical upgrade ($\Uparrow\phi$) makes all ϕ worlds, within aware cells of agents, more plausible than all $\neg\phi$ worlds, and within these two zones, the old ordering remains. I.e. $W^{\Uparrow\phi} = W$, $w \leqslant_a^{\Uparrow\phi} v$ iff $v \in ||\phi||_{\mathcal{M}}$ and $w \in ||\neg\phi||_{\mathcal{M}}$ or else if $w \leqslant_a v$, and $V^{\Uparrow\phi}(p) = V(p)$;

Conservative upgrade ($\uparrow\phi$) makes the best 'ϕ'-worlds, within the aware cells of agents, more plausible than all other worlds where ϕ is defined, while the old ordering on the rest of the worlds remains. I.e. $W^{\uparrow\phi} = W$, $w \leqslant_a^{\uparrow\phi} v$ iff either $v \in Max_{\leqslant_a}(||w||_a \cap ||\phi||_{\mathcal{M}})$ and $w \in (||\phi||_{\mathcal{M}} \cup ||\neg\phi||_{\mathcal{M}})$ or $w \leqslant_a v$, $V^{\uparrow\phi}(p) = V(p)$.

Finally, let us consider the event models for public and private announcements (see Section 2.2.2 on page 35). In fact, because the product update selects the worlds that do not falsify the precondition, setting the precondition to ϕ, like for DEL, already causes to only delete $\neg\phi$ world: both worlds in which ϕ is true and in which it is undefined are preserved. Furthermore, no postcondition needs to be defined because the old valuation is preserved for all worlds. I.e. the same event models for public and private announcements as defined for DEL can be applied to ParDEL+.

5.8 Raising Awareness with Event Models

Similar to announcements, we can capture the dynamics of raising awareness in an alternative way through event structures that specify how the agents observe the event. The obtained logic is called Partial Epistemic Action Logic (ParEAL), the adjustment to Epistemic Action Logic with partial valuation functions and weakly reflexive relations just like ParDEL(+) is to DEL. The syntax of ParEAL is analogous to that of EAL (Definition 2.25 on page 36), except that we consider *multipointed event models* $\langle \mathcal{E}, E^* \rangle$.

Definition 5.14 (Syntax of ParEAL)**.** Given a countable, non-empty set P of propositions and a finite, non-empty set $\mathcal{A}$ of agents, the *syntax*, $\mathcal{L}_{ParEAL}$, of (multi-agent) Partial Epistemic Action Logic (ParEAL) is defined in the following way.

$$\phi ::= p \mid \phi \wedge \psi \mid \neg\phi \mid K_a\phi \mid B_a\phi \mid [\langle \mathcal{E}, E^* \rangle]\phi$$

where $p \in P$ is a proposition, K_a and B_a are the knowledge and belief operators for each agent a, and $\langle \mathcal{E}, E^* \rangle$ are multipointed event models.

Frames and models of ParEAL are equivalent to ParDEL and ParDEL+ frames and models (Definitions 5.8 and 5.9 on pages 95, 96).

The multipointed event models for ParEAL are based on event models for DEL with postconditions (Definition 2.31 on page 41). These event models have introduced in [14] to accommodate for factual change. This is achieved through assigning to each event e and proposition p a substitution $post(e, p) \in \mathcal{L}_{DEL}$.

Concerning ParEAL, we consider a different definition of postconditions. Postconditions are considered functions that assign to each event e a *partial* function $post_e : P \to \{0, 1, \bot\}$. This partial function is then used to define the new valuation of p. This can likewise be used to define factual change, by letting $post_e(p)$ be equal to 0 or 1, or define the new valuation of p as 'undefined', when $post_e(p) = \bot$. The latter will be useful in Chapter 6 when we consider forgetting modalities.

Definition 5.15 (Event Model with Postconditions)**.** Let P be a countable, non-empty set of propositions and let $\mathcal{A}$ be a finite, non-empty set of agents. An *event model* for ParEAL is a quadruple $\mathcal{E} = \langle E, (R_a)_{a \in \mathcal{A}}, pre, post \rangle$ where

- E is a non-empty, finite set of *events*,
- $(R_a)_{a \in \mathcal{A}} \subseteq E \times E$ are the *accessibility relations* on E, one for each agent $a \in \mathcal{A}$,
- $pre : E \to \mathcal{L}_{ParEAL}$ is a *precondition function* assigning to each event a formula ϕ, and
- $post : E \to (P \to \{0, 1, \bot\})$ is a *postcondition function* assigning to each event a partial function $post_e : P \to \{0, 1, \bot\}$.

A *pointed event model (with postconditions)* is a pair $\langle \mathcal{E}, e \rangle$ where $\mathcal{E}$ is an event model with postconditions and $e \in E$.

A *multipointed event model (with postconditions)* is a pair $\langle \mathcal{E}, E^* \rangle$ where $\mathcal{E}$ is an event model with postconditions and $E^* \subseteq E$.

We will also write pre_e for $pre(e)$ and $post_e(p)$ for $(post(e))(p)$.

Multipointed event models $\langle \mathcal{E}, E^* \rangle$ with $E^* \subseteq E$ describe the sets of pointed event models $\langle \mathcal{E}, e \rangle$ with $e \in E^*$. When drawing a multipointed event model $\langle \mathcal{E}, E^* \rangle$, events are drawn as squares to distinguish them from ParEAL models and all $e \in E^*$ are double-squared to emphasize the points of reference. In the following, when given a multipointed event model $\langle \mathcal{E}, \{e_1, \ldots, e_n\} \rangle$, we will use E^* to denote $\{e_1, \ldots, e_n\}$. This E^* will be used to determine satisfiability for events.

Like for DEL, the product update of a ParEAL model $\mathcal{M}$ and an event model $\mathcal{E}$ determines what happens if an event takes place. In this product

update, the preconditions specify how to select the worlds in the product update. However, different to DEL, in ParEAL this selection takes place on the basis of 'not falsifying the precondition', instead of requiring that the precondition is verified. This is because lack of truth and falsity do not coincide on ParEAL. As a result, a precondition p selects the worlds in which either p is true, or p is undefined. That is, the worlds in which the precondition is undefined will remain too. This ensures that we can capture raising awareness in a natural way, as we will see.

Additionally, postconditions are used to specify the valuations that change in the product update. When the postcondition of a proposition is 1, this proposition becomes true, when it is 0, it becomes false, and when it is $\bot$, it becomes undefined. The latter will be used in Chapter 6 when we define forgetting modalities. Lastly, when the postcondition for a proposition is undefined, the old valuation is preserved.

Definition 5.16 (Product Update for ParEAL)**.** Let $\mathcal{M} = \langle W, (R_a^{\mathcal{M}})_{a\in\mathcal{A}}, V\rangle$ be a ParEAL model and $\mathcal{E} = \langle E, (R_a^{\mathcal{E}})_{a\in\mathcal{A}}, pre, post\rangle$ be an event model. Their *product update*, denoted by $\mathcal{M}\otimes\mathcal{E}$, is the triple $\langle W^{\mathcal{M}\otimes\mathcal{E}}, (R_a^{\mathcal{M}\otimes\mathcal{E}})_{a\in\mathcal{A}}, V^{\mathcal{M}\otimes\mathcal{E}}\rangle$ defined by:

- $W^{\mathcal{M}\otimes\mathcal{E}} = \{\langle w, e\rangle \in W \times E \mid \mathcal{M}, w \not\models pre_e\}$
- $\langle w, e\rangle R_a^{\mathcal{M}\otimes\mathcal{E}}\langle w', e'\rangle$ iff $\langle w, e\rangle, \langle w', e'\rangle \in W^{\mathcal{M}\otimes\mathcal{E}}$, $wR_a^{\mathcal{M}}w'$ and $eR_a^{\mathcal{E}}e'$
- $V^{\mathcal{M}\otimes\mathcal{E}}_{\langle w,e\rangle}(p) = \begin{cases} post_e(p) & \text{if } post_e(p) = 1 \text{ or } post_e(p) = 0 \\ \text{undefined} & \text{if } post_e(p) = \bot \\ V_w(p) & \text{otherwise} \end{cases}$

In the following we also refer to the events e such that $\mathcal{M}, w \not\models pre(e)$ as the events that *can be applied* to w. It follows from the definition that whenever no postcondition is defined, this means that the old valuation is preserved completely.

The event model for (public) raising awareness is as follows.

Definition 5.17 (Event Model for Raising Awareness)**.** The multipointed event model for *raising awareness* of a proposition p is $\langle \mathcal{E}_{+p}, \{e_p, e_{\bar{p}}\}\rangle$ where $\mathcal{E}_{+p} = \langle E_{+p}, (R_a)_{a\in\mathcal{A}}, pre, post\rangle$, with $E_{+p} = \{e_p, e_{\bar{p}}\}$, $R_a = \{e_p, e_{\bar{p}}\} \times \{e_p, e_{\bar{p}}\}$ and the pre- and postconditions defined as follows (see Figure 5.9):

- $pre_{e_p} = p$, $post_{e_p}(p) = 1$
- $pre_{e_{\bar{p}}} = \neg p$, $post_{e_{\bar{p}}}(p) = 0$

As mentioned, we also use E^*_{+p} to denote the points of reference $\{e_p, e_{\bar{p}}\}$.

Because the pre- and postconditions 'coincide' (that is, the valuation defined by the postcondition verifies the precondition), we can draw the event model as follows, where written p inside an event e means that $pre_e = p$ and $post_e(p) = 1$ and written $\neg p$ means that $pre_e = \neg p$ and $post_e(p) = 0$:

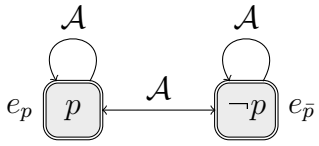

Figure 5.9: The event model $\mathcal{E}_{+p}$ for raising awareness of p, $+p$.

The event model for raising awareness then duplicates all the worlds of a model in which p was undefined, making p in one world and false in the other, while it preserves the other worlds. This is because worlds w such that $V_w(p)$ is undefined do not falsify p nor $\neg p$, hence $\mathcal{W}, w \not\models pre_{e_p}$ and $\mathcal{W}, w \not\models pre_{e_{\bar{p}}}$ and both events e_p and $e_{\bar{p}}$ can be applied to w. For an example of the application of $\mathcal{E}_{+p}$ to a model, see Figure 5.10.

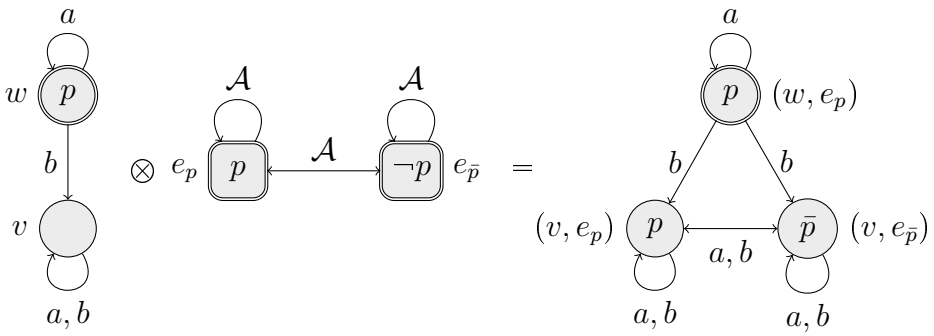

Figure 5.10: The event $\mathcal{E}_{+p}$ applied to the epistemic model on the left.

Satisfiability is defined analogously to events for EAL (Definition 2.28 on page 38), but now with respect to multipointed event models. The multiple points of reference are used to enforce that, whenever p was initially undefined at w of $\langle \mathcal{M}, w \rangle$ and awareness of p is raised, so that it is duplicated in two worlds $\langle w, e_p \rangle$ and $\langle w, e_{\bar{p}} \rangle$, both these duplicated worlds must make ϕ true in order for $[+p]\phi$ to hold at w. In this way, satisfiability for the event for raising awareness is equivalent to the clause for $+p$ in Definition 5.10 (page 96).

Definition 5.18 (Satisfiability for Multipointed Events)**.** Given a multipointed event model $\langle \mathcal{E}, E^* \rangle$, *satisfiability* for ParEAL with event models extends satisfiability of ParDEL (Definition 5.10 on page 96) by:

$$\mathcal{M}, w \vDash [\langle \mathcal{E}, E^* \rangle]\phi \text{ iff } \forall e \in E^* : \mathcal{M}, w \neq pre(e) \text{ implies } \mathcal{M} \otimes \mathcal{E}, \langle w, e \rangle \vDash \phi$$
$$\mathcal{M}, w \dashv [\langle \mathcal{E}, E^* \rangle]\phi \text{ iff } \exists e \in E^* : \mathcal{M}, w \neq pre(e) \text{ and } \mathcal{M} \otimes \mathcal{E}, \langle w, e \rangle \dashv \phi$$

Figures 5.8 (page 101) and 5.10 illustrate that $+p$ as defined as a model transformer in Definition 5.12 (page 99) and the multipointed event model $\mathcal{E}_{+p}$ lead to the same model.

Proposition 5.5. For any ParDEL+/ParEAL model $\mathcal{M}$ and any $p \in P$, $\mathcal{M}^{+p}$ is bisimilar to $\mathcal{M} \otimes \mathcal{E}_{+p}$.

Proof. $\mathcal{M}^{+p}$ is $\mathcal{M}$ with every world $w \in \mathcal{M}$ such that $p \notin Dom(V_w)$ duplicated into $\langle w, 0 \rangle$ and $\langle w, 1 \rangle$ making p false and true, respectively, while the relations and valuations for other propositions remain the same. This is exactly the same as $\mathcal{M} \otimes \mathcal{E}_{+p}$, by replacing 1 by e_p and 0 by $e_{\bar{p}}$. I.e. the same valuations and the same relations hold. Therefore Z defined as $\forall \langle w, i \rangle \in W^{+p}$: $\langle w, 0 \rangle Z \langle w, e_{\bar{p}} \rangle$ and $\langle w, 1 \rangle Z \langle w, e_p \rangle$ is a bisimulation. □

Furthermore, the semantics is equivalent for $+p$ and $\langle \mathcal{E}_{+p}, \{e_p, e_{\bar{p}}\} \rangle$. Therefore, we can use these two interchangeably.

Proposition 5.6. For any ParDEL+/ParEAL model $\mathcal{M}$, any world $w \in \mathcal{M}$, any proposition $p \in P$ and any formula ϕ, it holds that:

$$\mathcal{M}, w \vDash [+p]\phi \text{ iff } \mathcal{M}, w \vDash [\langle \mathcal{E}_{+p}, \{e_p, e_{\bar{p}}\} \rangle]\phi$$
$$\mathcal{M}, w \dashv [+p]\phi \text{ iff } \mathcal{M}, w \dashv [\langle \mathcal{E}_{+p}, \{e_p, e_{\bar{p}}\} \rangle]\phi$$

Proof. We prove the case of verification, the case of falsification is analogous. It holds that $\mathcal{M}, w \vDash [+p]\phi$ iff $\forall \langle w, i \rangle \in W^{+p}$ with $i \in \{0, 1\}$: $\mathcal{M}^{+p}, \langle w, i \rangle \vDash \phi$. But since $\mathcal{M}^{+p}$ and $\mathcal{M} \otimes \mathcal{E}_{+p}$ are bisimilar (Proposition 5.5), they can be exchanged so that, by renaming $\langle w, i \rangle$ with events $e_p, e_{\bar{p}}$, $\mathcal{M}, w \vDash [+p]\phi$ iff $\forall \langle w, e \rangle \in W^{\mathcal{M} \otimes \mathcal{E}}$ with $e \in \{e_p, e_{\bar{p}}\}$: $\mathcal{M} \otimes \mathcal{E}, \langle w, e \rangle \vDash \phi$. Therefore, $\mathcal{M}, w \vDash [+p]\phi$ iff $\mathcal{M}, w \vDash [\langle \mathcal{E}_{+p}, \{e_p, e_{\bar{p}}\} \rangle]\phi$. □

5.8.1 Private Raising Awareness

When defining event models, more complex upgrades can be visualized. For example, private announcements have been discussed in Section 2.2.2

(page 35). Similarly, we here consider the dynamics of raising *private* awareness: a group G of agents raises their awareness of a proposition p, and this occurs in full privacy, i.e. the other agents ($\mathcal{A}\backslash G$) do not observe the upgrade or even consider that it took place.

The event model for raising private awareness looks similar to that for public awareness (Definition 5.17 on page 104), but instead of a single event where the proposition p is true and another event with precondition $\neg p$, we consider three events: one where the precondition is p, one where the precondition is $\neg p$ and a third where the precondition is $\top$. The first two events duplicate the worlds in which p is undefined, and the last event ensures that the other agents do not observe the event. This is analogous to the event model for private announcements (Definition 2.30 on page 39).

Definition 5.19 (Event Model for Private Raising Awareness)**.** The multipointed event model for *private raising awareness* of a proposition p amongst a group $G \subseteq \mathcal{A}$ of agents is $\langle \mathcal{E}_{+Gp}, \{e_p, e_{\bar{p}}\}\rangle$ where $\mathcal{E}_{+Gp} = \langle E_{+Gp}(R_a)_{a\in\mathcal{A}}, pre\rangle$ where $E_{+Gp} = \{e_p, e_{\bar{p}}, e_\top\}$, $R_a = (\{e_p, e_{\bar{p}}\} \times \{e_p, e_{\bar{p}}\}) \cup \{\langle e_\top, e_\top\rangle\}$ for $a \in G$ and $R_a^{\mathcal{E}} = \{e_p, e_{\bar{p}}, e_\top\} \times \{e_\top\}$ otherwise, and the pre- and postconditions are defined as follows (see Figure 5.11):

- $pre_{e_p} = p$, $post_{e_p}(p) = 1$
- $pre_{e_{\bar{p}}} = \neg p$, $post_{e_{\bar{p}}}(p) = 0$
- $pre_{e_\top} = \top$

As before, we also use E^*_{+Gp} to denote the points of reference $\{e_p, e_{\bar{p}}\}$. We also refer to the multipointed event model $\langle \mathcal{E}_{+Gp}, \{e_p, e_{\bar{p}}\}\rangle$ as $+_G p$. The multipointed event model does not include $e_\top$. This is because $e_\top$ only represents that the other agents, not belonging to G, do not observe the event, but it is not what actually takes place: namely that the group G does raise their awareness.

Private raising awareness could be used to raise awareness of a single agent when $G = \{a\}$, i.e. $+_{\{a\}}p$. In that case, only agent a raises her awareness of p while the other agents remain in the old state.

Because the pre- and postconditions 'coincide' (that is, the valuation of the postcondition verifies the precondition), we can draw the event model as in Figure 5.11, where written p inside an event e means that $pre(e) = p$ and $post_e(p) = 1$, and $\top$ denotes $pre_e = \top$ and $post_e(p)$ is undefined for every $p \in P$.

Like with public and private announcements, privately raising the awareness of p for the group of all agents, $G = \mathcal{A}$, amounts to raising (public) awareness of p.

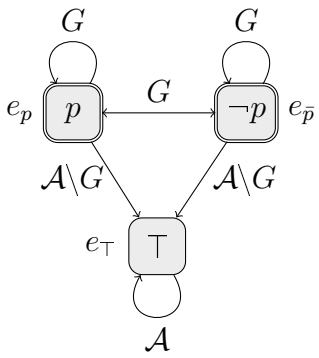

Figure 5.11: The event model $\mathcal{E}_{+_G p}$ for private raising awareness of p, $+_G p$.

Proposition 5.7. Given a pointed ParEAL model $\langle \mathcal{M}, w \rangle$, a proposition p and a formula ϕ. Then $\mathcal{M}, w \vDash [+_{\mathcal{A}} p]\phi$ if and only if $\mathcal{M}, w \vDash [+p]\phi$ and $\mathcal{M}, w \dashv [+_{\mathcal{A}} p]\phi$ if and only if $\mathcal{M}, w \dashv [+p]\phi$.

Proof. The model $\mathcal{M}^{+_{\mathcal{A}} p}$ is made of two disconnected components: the first one is exactly $\mathcal{M}^{+p}$ resulting from the product of $\mathcal{M}$ with the part of $\mathcal{E}_{+_{\mathcal{A}} p}$ which is identical to $\mathcal{E}_{+p}$ (i.e. the events e_p and $e_{\bar{p}}$ and the relations between them), the second component results from the product of $\mathcal{M}$ with $e_\top$. These are disconnected because the edges $\mathcal{A}\backslash G$ are empty when $G = \mathcal{A}$.

[only if] Assume that $\mathcal{M}, w \vDash [+p]\phi$. Then $\forall \langle w, e \rangle \in W^{+p}$ s.t. $e \in E^*_{+p} = \{e_p, e_{\bar{p}}\}$: $\mathcal{M}^{+p}, \langle w, e \rangle \vDash \phi$. But these worlds are exactly the worlds in the image of w under $+_{\mathcal{A}} p$ that are used to determine satisfiability, and these worlds satisfy the same conditions. Therefore $\mathcal{M}, w \vDash [+_{\mathcal{A}} p]\phi$.

Similarly, assume that $\mathcal{M}, w \dashv [+p]\phi$. Then $\exists \langle w, e \rangle \in W^{+p}$ s.t. $e \in E^*_{+p} = \{e_p, e_{\bar{p}}\}$: $\mathcal{M}^{+p}, \langle w, e \rangle \dashv \phi$. But again, this world is also in the image of w under $+_{\mathcal{A}} p$ that is used to determine satisfiability, and this world satisfies the same conditions. Hence $\mathcal{M}, w \dashv [+_{\mathcal{A}} p]\phi$.

[if] We can reverse the reasoning above to show that $\mathcal{M}, w \vDash [+_{\mathcal{A}} p]\phi$ implies $\mathcal{M}, w \vDash [+p]\phi$ and $\mathcal{M}, w \dashv [+_{\mathcal{A}} p]\phi$ implies $\mathcal{M}, w \dashv [+p]\phi$ because satisfiability for $+_{\mathcal{A}} p$ is determined by the worlds $\langle w, e_p \rangle$ and $\langle w, e_{\bar{p}} \rangle$ in the image of w under $+_{\mathcal{A}} p$, because $e_\top \notin E^*_{+_{\mathcal{A}} p}$. And these worlds are also in the image of w under $+p$ and again, satisfy the same conditions. So whenever something holds for all these worlds (for $\vDash$) or for at least one of the worlds ($\dashv$) under $+_{\mathcal{A}} p$, this must also be the case for these worlds under $+p$. □

Needless to say, private raising awareness for the group of all agents, i.e. $G = \mathcal{A}$, is not equivalent to privately raising awareness for agent a_1, then a_2, then a_3, ..., until a_n. I.e., satisfiability for $+_{\{a_1\}} p; \ldots; +_{\{a_n\}} p$ is not the

same as satisfiability for $+_G p$. This is because in that case, neither agent observes the raising awareness operation for the other agents. Hence, even though each agent becomes aware of p, this is not common information – whereas it was for the group G with the upgrade $+_G p$. More concretely, the formula $[+p]K_a K_b(p \vee \neg p)$ is valid on all ParDEL+/ParEAL models, but $[+_b p; +_a p]K_a K_b(p \vee \neg p)$ is not.

In Figure 5.12, the event model for private raising awareness for a single agent is applied to a pointed epistemic model. This situation is clearly different from the case in which this agent raised her awareness publicly, as depicted in Figure 5.10, because the agents do not have common awareness of p.

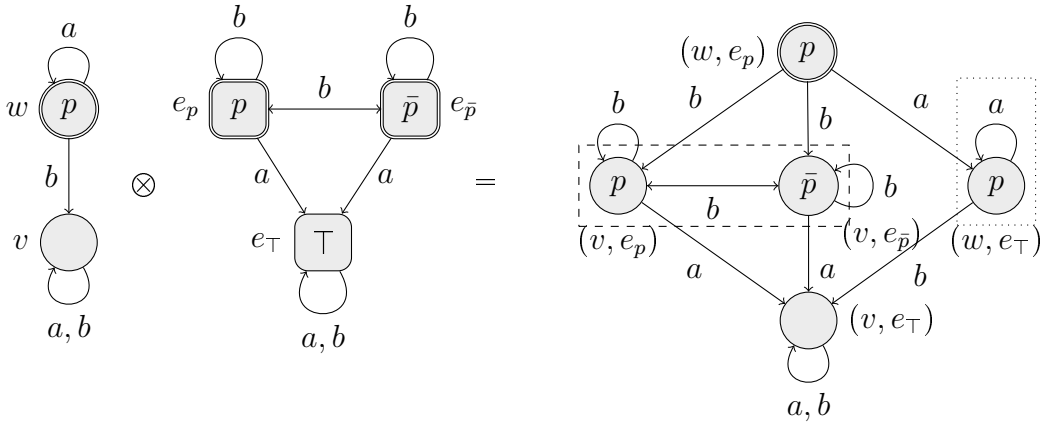

Figure 5.12: The event model of $+_{\{b\}} p$, $\mathcal{E}_{+_{\{b\}}p}$, applied to an epistemic model $\mathcal{M}$, on the left. In the product update (right), the aware cell $||(w, e_p)||_a$ is indicated by a dotted box and the aware cell $||(w, e_p)||_b$ by a dashed box.

5.9 Raising Awareness without Disclosing Truth

By switching to partial valuations, raising awareness functions via extending the valuation function. This implies that, unlike previous work on awareness where raising awareness also unveiled truth values [15, 38, 51], raising awareness does not disclose truth. In this section, this is proved.

Raising the awareness of agents does not lead them to acquire new knowledge or beliefs. Hence:

- everything agents previously knew or believed, is still known and believed after raising awareness (*knowledge and belief preservation*), and

- everything* known or believed after raising awareness, was already known or believed before (*knowledge and belief correspondence*).

In the second clause, there is a catch. This holds only for formulas not including the proposition awareness is raised of. Hence for p-free ϕ when the modality is $+_G p$. This is for the obvious reason that tautologies involving p, for example $p \vee \neg p$, or formulas of the form $p \to \phi$ where ϕ was previously known or believed, will also be known or believed after raising the awareness of p. The term 'everything*' is used to exclude these cases.

In fact, preservation and correspondence do not only apply to epistemic formulas but also to non-epistemic formulas: truth is also preserved and corresponds in the case of formulas not including the proposition awareness is raised of. This is straightforward: raising awareness extends the valuation function with proposition p but does not alter the valuation of other propositions. Therefore, formulas not involving p keep their old truth value, and formulas involving p acquire a truth value.

Proposition 5.8 (Truth preservation and correspondence)**.** Let $\mathcal{M}$ be a ParEAL model and let w be a world in $\mathcal{M}$. Then for any group $G \subseteq \mathcal{A}$ and for any formula ϕ:

$$\mathcal{M}, w \vDash \phi \to [+_G p]\phi \qquad \text{(truth preservation)}$$

and for any ψ not containing p:

$$\mathcal{M}, w \vDash [+_G p]\psi \to \psi \qquad \text{(truth correspondence)}$$

Proof. For preservation, assume that $\mathcal{M}, w \vDash \phi$. Then, after raising the awareness of p, for all $\langle w, e\rangle \in W^{+_G p}$ with $e \in E^*_{+_G p}$ the valuation $V^{+_G p}_{\langle w,e\rangle}$ is either equal to V_w (in case $p \in Dom(V_w)$) or extends it V_w by a valuation for proposition p (in case $p \notin Dom(V_w)$). In both cases, for all $q \in Dom(V_w)$: $V^{+_G p}_{\langle w,e\rangle}(q) = V_w(q)$. Therefore, for all $\langle w, e\rangle \in W^{+_G p}$ with $e \in E^*_{+_G p}$ it holds that $\mathcal{M}^{+_G p}, \langle w, e\rangle \vDash \phi$, and thus $\mathcal{M}, w \vDash [+_G p]\phi$.

For correspondence, assume that $\mathcal{M}, w \vDash [+_G p]\psi$ for some formula ψ not containing p. This means that $\mathcal{M}^{+_G p}, \langle w, e\rangle \vDash \psi$ for all $\langle w, e\rangle \in W^{+_G p}$ with $e \in E^*_{+_G p}$. But since ψ did not contain p, for all q in ψ: $V^{+_G p}_{\langle w,e\rangle}(q) = V_w(q)$. Therefore, it must also hold that $\mathcal{M}, w \vDash \psi$. □

In the proof of Proposition 5.8, there is no distinction between propositional formulas ϕ and non-propositional formulas ϕ, for example $\phi = K_a\psi$, because in both cases all the propositions q occuring in ϕ are evaluated at v. For the non-propositional formulas, this is true because of the specification property for valuations in ParEAL: this causes any propositions q occuring

in ϕ such that $K_a\phi$, even if $\phi = K_b\psi$, to have a truth value at v. Then, the fact that raising awareness does not alter the valuation of these propositions q gives rise to the proof.

Furthermore, raising the awareness of p does not disclose the truth of p itself. That is, whenever p was undefined ($\not\models p$ and $\not\models \neg p$), it is false that after awareness is raised of p, p is true, and that p is false, i.e. $\not\models [+_G p]p$ and $\not\models [+_G p]\neg p$.

Proposition 5.9 (Raising awareness without disclosing truth)**.** Let $\mathcal{M}$ be a ParDEL+/ParEAL model and let w be a world in $\mathcal{M}$. Then for any group $G \subseteq \mathcal{A}$ and any proposition p:

$$\text{If } \mathcal{M}, w \not\models p \text{ and } \mathcal{M}, w \not\models \neg p \text{ then } \mathcal{M}, w \not\models [+_G p]p \text{ and } \mathcal{M}, w \not\models [+_G p]\neg p$$

Proof. Whenever $\mathcal{M}, w \not\models p$ and $\mathcal{M}, w \not\models \neg p$, it means that $p \notin Dom(V_w)$. Hence, $V^{+_G p}_{\langle w, e_p\rangle}(p) = 1$ and $V^{+_G p}_{\langle w, e_{\bar{p}}\rangle}(p) = 0$. Thus $\mathcal{M}^{+_G p}, \langle w, e_p\rangle \models p$ and $\mathcal{M}^{+_G p}, \langle w, e_{\bar{p}}\rangle \not\models p$. By satisfiability for ParDEL+, then $\mathcal{M}^{+_G p}, \langle w, e_p\rangle \not\models \neg p$. Thus, since $E^*_{+_G p} = \{e_p, e_{\bar{p}}\}$, $\mathcal{M}, w \not\models [+_G p]p$ and $\mathcal{M}, w \not\models [+_G p]\neg p$. □

In conclusion, raising awareness does not disclose truth values nor add new truths, other than the trivial cases ($p \vee \neg p$ or $p \to \phi$, where ϕ was true before, amongst others). This means that the raising awareness modalities introduced for ParEAL truly disconnect awareness from truth.

5.10 Discussion

In this chapter, we introduced awareness through partial valuation functions and weakly reflexive relations and we showed that raising awareness modalities can be defined that raise awareness without disclosing truth values. This is useful to model communication and interactions between agents that use *different* and *dynamic* vocabularies, and therefore provides a good alternative for DEOL to capture the Alignment Repair Game (ARG).

Yet, ARG is not the only use of the logic introduced. The semantics for ParDEL+ and ParEAL do not only provide a more general framework to model situations not restricted to full vocabulary awareness, ParEAL also provides a way to interpret complex dynamic upgrades such as private announcements. As briefly discussed in Section 2.2.2 (page 35), private announcements, which have been introduced for DEL via event models, are typically not permitted in DEL models because they do not preserve their properties. In particular, they violate reflexivity of the accessibility relations of agents. This is because for any $a \notin G$, the private announcement $!_G\phi$ does

not have a reflexive arrow for a at e_ϕ (Definition 2.30 on page 39). Hence, there is also no reflexive arrow for agent a at any world $\langle w, e_\phi \rangle$ in the product update, causing the knowledge of this agent to no longer be factive.

Consequently, this means that the only events allowed to take place are those that preserve the properties of the relations of the models, i.e. those that satisfy the same properties, in particular reflexivity. For a framework as rich as DEL, this clearly is a pity. It means that not only private announcements are not permitted, but *any* upgrade involving a degree of privacy is not. Because for any form of privacy, reflexivity needs to be dropped.

Instead of restricting the events that can take place to preserve the properties, another solution would be to consider a logic with different properties. Of course, this can go very far and it is not desirable to drop all the properties of the logic. However, ParEAL, which replaces reflexivity by weak reflexivity, drops 'enough' to allow for private events. Indeed, private announcements do preserve the properties of ParEAL models: in particular, weak reflexivity is preserved as long as the event models satisfy weak reflexivity - which is the case for the private upgrades we considered in this Chapter. This means that ParEAL could be used as a framework for private communication between agents. As such, already without awareness, ParEAL is a more general framework to model communication and knowledge or belief change.

Dropping reflexivity also has a drawback: factivity no longer holds. However, it still holds locally within the aware cells of agents. This corresponds to the weak reflexivity condition required for the accessibility relations. As a result, knowledge may be considered 'intermediate' between $S5$ (within aware cells) and $KD45$ (outside aware cells). It is an open question to what axiom such an intermediate knowledge operator corresponds and that formalizes weak reflexivity as a condition on the models.

5.11 Conclusion

ParDEL+/ParEAL formalizes a notion of awareness based on partial valuation functions and weakly reflexive relations. Raising awareness and private raising awareness modalities have been discussed and it was shown that they do not disclose truth. This allows agents to use different and dynamic vocabularies to represent their knowledge and beliefs. For this reason, these logics are a good candidate to improve the logical model of ARG: heterogeneity is ensured via awareness and agents can learn new vocabulary, without learning its truth, on the fly via the raising awareness modalities. This can be used to reduce the differences between adaptive agents and logical agents concerning their vocabulary awareness.

In Chapter 6, awareness will be used to address another difference between the two types of agents via defining modalities for forgetting. Then, in Chapter 7, awareness will be taken back to ARG to re-examine the formal properties of the adaptation operators. This will answer whether or not introducing awareness closer resembles ARG than DEOL.

Chapter 6

Forgetting

> I've a grand memory for forgetting.
>
> Robert Louis Stevenson

In Chapter 4, one of the differences between adaptive and logical agents discussed is the ability of adaptive agents to forget cases and to focus on general knowledge, while logical agents cannot forget. With a formal notion of awareness, we can also define forgetting modalities.

Adaptive agents, who play ARG, discard the classification of object announced and focus on improving the alignments, which is sufficient for them to prevent the same failure from occurring in the future and converge to successful communication. Therefore, since they do not need specific examples to communicate successfully, they do not store them. This means that in the translation of this state, $K_a(C_b(o))$ does not hold, e.g. $\tau(\alpha(s)) \not\models K_a(C_b(o))$ in Figure 6.1. On the contrary, in the logical model of ARG in DEOL, agents cannot discard or forget these classifications of objects in the same way. This is inherent to the definition of Dynamic Epistemic Logic (DEL), on which Dynamic Epistemic Ontology Logic (DEOL) was based that was used to translate ARG. After the announcement $!C_b(o)$ in the second step of ARG (Definition 2.12 on page 20), both agents know this, in particular agent a. Therefore $\tau(s)^{\delta(\alpha)} \models K_a(C_b(o))$, see Figure 6.1. Moreover, DEOL, like DEL, does not have ways for agents to 'unlearn' what they know. Therefore, whatever upgrade is applied after, agent a still knows that $C_b(o)$.

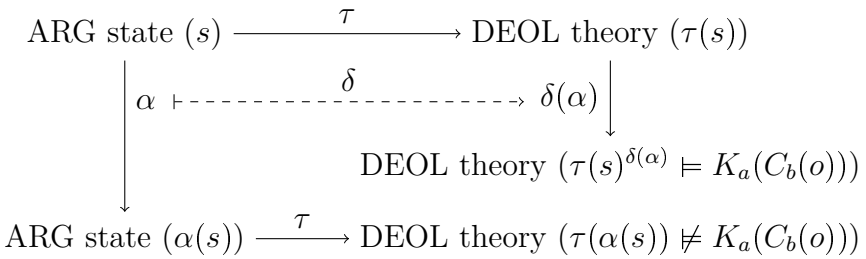

Figure 6.1: The translation from ARG states (s) to DEOL theories (τ) and from adaptation operators (α) to dynamic upgrades (δ), where $K_a(C_b(o))$ can be deduced from $\tau(s)^{\delta(\alpha)}$ but not from $\tau(\alpha(s))$.

This difference in the ability to remember or forget is present in the proof for incompleteness of the adaptation operators (Proposition 3.8 on page 70), from which we can conclude that there is more to be deduced than adaptive agents do. In this chapter, instead, we want to bring the logical model yet closer to ARG and introduce forgetting modalities.

In the following, first two different ways to forget are discussed and whether these can serve to reverse the raising awareness modality. Then, these are formalized. This lays the ground for Chapter 7 where we will bring awareness and forgetting back to ARG and define a new translation.

6.1 Two Types of Forgetting

Raising awareness as defined in Definition 5.12 (page 99) is a modality that transforms unaware agents into *uncertain*, or ignorant, agents. This means that agents, who were previously unaware of p, become aware of p after raising awareness of it, but do not come to know or believe the truth value: they consider it equally likely that p is true as that it is false. Formally, raising awareness duplicates each world in which the proposition awareness is raised of is undefined, making it true in one world and false in the other. Relations from and to the worlds are preserved, and so are the valuations of other propositions.

To forget, we consider two options that were introduced in [41]:

- *Becoming unaware*: Agents become unaware of a fact, even if they knew or believed it before.
- *Becoming uncertain*: Agents forget the truth value of a fact, but are "embarrassingly aware of [their] current ignorance" [41], i.e. they know that they do not know p after forgetting it.

The first of these, 'becoming unaware', seems a natural 'reverse' modality for raising awareness.

6.1.1 Becoming Unaware $-p$

To forget, one could simply delete the valuations of the proposition p that is to be forgotten, possibly merging worlds up to bisimilarity. Such an operation makes agents, whatever they knew or believed about p, unaware of p. It will be referred to as $-p$, or *forgetting awareness*. While aware agents become unaware through $-p$, unaware agents are not affected by it, see Figure 6.2.

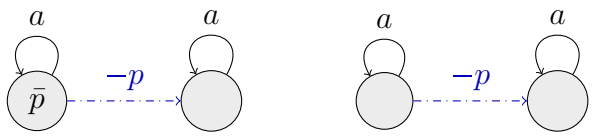

Figure 6.2: Forgetting p as a 'becoming unaware' modality, $-p$, applied to a world in which p is defined (left) and in which p is undefined (right). The modality is indicated by blue dashdotted arrows.

This type of forgetting is of the type 'becoming unaware'. It reverses the raising awareness modality on models in which the proposition is initially undefined and raising awareness and forgetting are applied subsequently, see Figure 6.3.

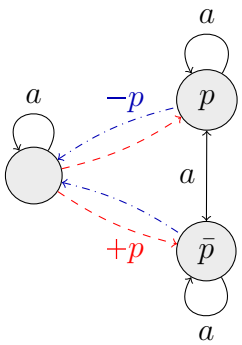

Figure 6.3: Raising the awareness of p, $+p$ (left to right, indicated by red dashed arrows), and forgetting p, $-p$ (right to left, indicated by blue dashdotted arrows), merging worlds up to bi-similarity, for an agent that is unaware of p.

However, it is not generally a reverse of raising awareness. Consider Figure 6.4, in which p is defined in the initial model. In this case, raising awareness does not alter the model, while forgetting awareness does: all valuations of p are deleted, which does not correspond to the initial model. This is because, while raising awareness does not affect worlds in which p is defined, forgetting awareness does affect these worlds. Hence, performing $+p$ and subsequently $-p$ does not lead back to the initial situation but instead leads to the model in which p is undefined.

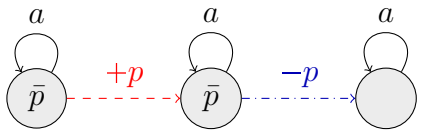

Figure 6.4: Raising awareness of p, $+p$ (indicated by red dashed arrows), and forgetting p, $-p$ (indicated by blue dashdotted arrows), for an agent that initially knows that p is false.

Moreover, different from raising awareness, forgetting awareness cannot be not disconnected from forgetting truth. When agents forget awareness and become unaware, so do they forget the truth value. Recall that the raising awareness modality as defined in Chapter 5 raised awareness *without* disclosing truth to the agents: truth is preserved and, for formulas without occurrences of the proposition that awareness is raised of, corresponds (Proposition 5.8 on page 110). This is not true for a forgetting modality that deletes valuations because forgetting awareness cannot be performed independently from 'becoming uncertain': through forgetting awareness of a proposition p, whatever was known or believed about p will no longer be known or believed.

6.1.2 Becoming Uncertain $\ominus p$

The other approach to forgetting, the type 'becoming uncertain', requires a different definition. Instead of deleting the valuations of a proposition p, we want that agents solely drop the truth value of p, without losing awareness. This can be achieved through 'copying' all the worlds in which p is defined and 'flip' the truth value in the copied world, see Figure 6.5. This means that, if in a world p is false, forgetting copies this world to a world in which the relations and valuations are preserved, except the truth value of p – hence, in this copied world, p is true. This will be referred to as $\ominus p$, or *forgetting*

truth. Again, like for forgetting awareness, whenever p was already undefined in a world, $\ominus p$ does not affect it, see Figure 6.5.

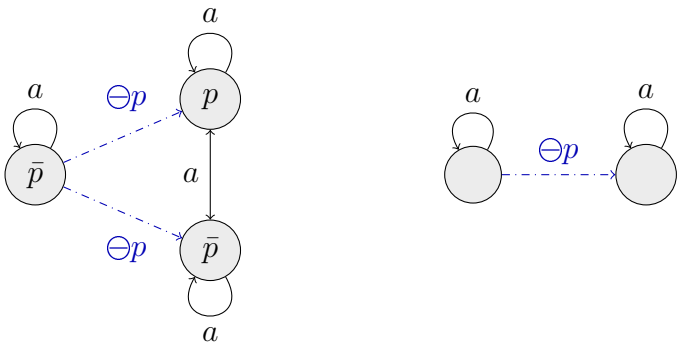

Figure 6.5: Forgetting p as a 'becoming uncertain' modality, $\ominus p$, applied to a world in which agent a knows that p is false (left) and in which agent a in aware of p (right). The modality is indicated by blue dashdotted arrows.

After forgetting truth, agents remain aware of the proposition, but become uncertain about it. Therefore, it is surely not a reverse of raising awareness: it preserves awareness and deletes truth. Rather, it may be considered as a reverse for announcements, on the condition that before announcing p, agents were uncertain about p, see Figure 6.6. In this case, it does not matter whether the announcement was $!p$ or $!\neg p$, as long as it caused agents from being uncertain to knowing the truth value of p.

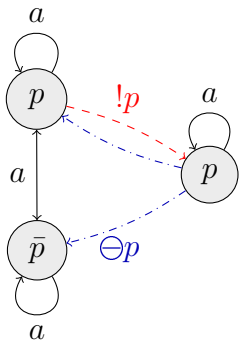

Figure 6.6: Announcing p, $!p$ (left to right, indicated by red dashed arrows), and forgetting p, $\ominus p$ (right to left, indicated by blue dashdotted arrows).

However, like $-p$ is not a true reverse of $+p$, $\ominus p$ is not a true reverse of

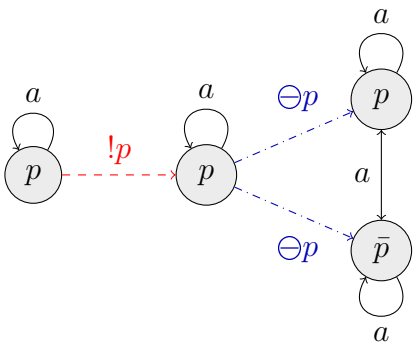

Figure 6.7: Announcing p, $!p$ (indicated by red dashed arrows), and forgetting the truth value of p, $\ominus p$ (indicated by blue dashdotted arrows), when p was initially true.

$!p$. This is especially true if p was already true (or false) in the model, see Figure 6.7. In that case, the announcement $!p$ does not affect it, while $\ominus p$ does by copying all the p-worlds and changing the valuation of p in these worlds.

6.2 Definitions of Forgetting

We now formalize the two types of forgetting, using the extension of Partial Dynamic Epistemic Logic with both raising awareness and forgetting. We call this logic ParDEL$\odot$.

Definition 6.1 (Syntax of ParDEL$\odot$)**.** Given a countable, non-empty set P of propositions and a finite, non-empty set $\mathcal{A}$ of agents, the *syntax*, $\mathcal{L}_{ParDEL\odot}$ of (multi-agent) Partial Dynamic Epistemic Logic with raising awareness and forgetting (ParDEL$\odot$) is defined in the following way.

$$\phi ::= p \mid \phi \wedge \psi \mid \neg\phi \mid K_a\phi \mid B_a\phi \mid [\dagger\phi]\psi \mid [\odot p]\phi$$

where p is a proposition, K_a and B_a are the knowledge and belief operators for each agent a, and $\dagger\phi$ with $\dagger \in \{!, \Uparrow, \uparrow\}$ and $\odot p$ with $\odot \in \{+, -, \ominus\}$ the dynamic upgrades.

Frames and models of ParDEL$\odot$ are equivalent to frames and models of ParDEL, ParDEL+ and ParEAL (Definitions 5.8 and 5.9 on pages 95, 96).

Forgetting awareness, $-p$, deletes the valuation of the proposition that is forgotten of. More precisely, it reduces the scope of the valuation function by p.

Definition 6.2 (Forgetting Awareness $(-p)$)**.** Let $\mathcal{M} = \langle W, (R_a)_{a\in\mathcal{A}}, V\rangle$ be a ParDEL$\odot$ model and let $p \in P$. Then *forgetting awareness* $-p$ is a model transformer $-p : \mathcal{M} \mapsto \mathcal{M}^{-p}$ where $\mathcal{M}^{-p}$ is the triple $\langle W^{-p}, (R_a^{-p})_{a\in\mathcal{A}}, V^{-p}\rangle$ defined by:

- $W^{-p} = W$
- $wR_a^{-p}v$ iff wR_av
- $V_w^{-p}(q) = \begin{cases} V_w(q) & \text{if } q \neq p \\ \text{undefined} & \text{otherwise} \end{cases}$

One may choose to merge worlds up to bisimilarity after applying the forgetting modality.

Unlike raising awareness, we here write $p \in P$ because forgetting awareness, or truth, is not interesting for other cases, whereas this was for raising awareness: in this way, the set of propositions could be extended.

Forgetting truth, $\ominus p$, copies the worlds in which p is true or false, and 'flips' the truth value. In other words, it changes the valuation of p in these worlds.

Definition 6.3 (Forgetting Truth $(\ominus p)$)**.** Let $\mathcal{M} = \langle W, (R_a)_{a\in\mathcal{A}}, V\rangle$ be a ParDEL$\odot$ model and let $p \in P$. Then *forgetting truth* $\ominus p$ is a model transformer $\ominus p : \mathcal{M} \to \mathcal{M}^{\ominus p}$ where $\mathcal{M}^{\ominus p}$ is the triple $\langle W^{\ominus p}, (R_a^{\ominus p})_{a\in\mathcal{A}}, V^{\ominus p}\rangle$ defined by:

- $W^{\ominus p} = (W|_p \cup W|_{\neg p}) \times \{0,1\} \cup W\backslash(W|_p \cup W|_{\neg p}) \times \{0\}$
- $\langle w,i\rangle R_a^{\ominus p}\langle v,j\rangle$ iff wR_av
- $V^{\ominus p}_{\langle w,i\rangle}(q) = \begin{cases} V_w(q) & \text{if } q \neq p \\ V_w(p) & \text{if } q = p \text{ and } w \in W\backslash(W|_p \cup W|_{\neg p}) \\ i & \text{otherwise} \end{cases}$

Satisfiability for ParDEL$\odot$ extends that of ParDEL+ in the natural way.

Definition 6.4 (Satisfiability for ParDEL$\odot$)**.** *Satisfiability* for ParDEL$\odot$ extends that of ParDEL+ (Definition 5.13 on page 99) by:

$$\begin{array}{ll} \mathcal{M}, w \vDash [-p]\phi & \text{iff } \mathcal{M}^{-p}, w \vDash \phi \\ \mathcal{M}, w \vDash [\ominus p]\phi & \text{iff } \forall\langle w,i\rangle \in W^{\ominus p} : \mathcal{M}^{\ominus p}, \langle w,i\rangle \vDash \phi \end{array}$$

for verification ($\vDash$) and for falsification ($\dashv$):

$$\begin{array}{ll} \mathcal{M}, w \dashv [-p]\phi & \text{iff } \mathcal{M}^{-p}, w \dashv \phi \\ \mathcal{M}, w \dashv [\ominus p]\phi & \text{iff } \exists\langle w,i\rangle \in W^{\ominus p} : \mathcal{M}^{\ominus p}, \langle w,i\rangle \dashv \phi \end{array}$$

6.2.1 Forgetting with Event Models

We can also capture the forgetting modalities with event models as defined for Partial Epistemic Action Logic (ParEAL). This enables to also define private forgetting, analogous to private announcements and private raising awareness. This will be useful when we return to the logical model of ARG.

We follow the definitions of event models, product updates and satisfiability as stipulated in Section 5.8 (page 102) for ParEAL. Recall that a precondition is a function $pre : E \to \mathcal{L}_{ParEAL}$ that determines which worlds appear in the product update (those such that $\mathcal{M}, w \not\models pre(e)$), and a postcondition is a partial function $post : E \to (P \to \{0, 1, \bot\})$ determining the new valuation function (Definition 5.15 on page 103):

$$V^{\mathcal{M}\otimes\mathcal{E}}_{\langle w,e\rangle}(p) = \begin{cases} post_e(p) & \text{if } post_e(p) = 1 \text{ or } post_e(p) = 0 \\ \text{undefined} & \text{if } post_e(p) = \bot \\ V_w(p) & \text{otherwise} \end{cases}$$

Therefore, a postcondition 1 or 0 changes the valuation of p to true or false, a postcondition $\bot$ makes p undefined and if the postcondition is not defined for a proposition, its old valuation (true, false or undefined) is preserved. For raising awareness, only postconditions 1 and 0 were considered, but for forgetting, the option $\bot$ becomes useful to indicate that a valuation is 'deleted' from the model.

Forgetting awareness

First, let us look at the event model for forgetting awareness. To forget awareness, all worlds and relations are preserved, only the valuation is adjusted. In particular, forgetting awareness of p causes the valuation of p to be deleted, while preserving other valuations. The event model for forgetting awareness therefore consists of one event $e_{\not p}$ with precondition $\top$ and postcondition assigning p to $\bot$.

Definition 6.5 (Event Model for Forgetting Awareness)**.** The multipointed event model for *forgetting awareness* of a proposition p is $\langle \mathcal{E}_{-p}, \{e_{\not p}\}\rangle$ where $\mathcal{E}_{-p} = \langle E_{-p}, (R_a)_{a\in\mathcal{A}}, pre, post\rangle$, with $E_{-p} = \{e_{\not p}\}$, $R_a = \{\langle e_{\not p}, e_{\not p}\rangle\}$, $pre(e_{\not p}) = \top$ and $post_{e_{\not p}}(p) = \bot$ (see Figure 6.8).

As before, we also use E^*_{-p} to denote the point of reference $\{e_{\not p}\}$.

When drawing event models, we write ϕ on the first line of an event e to indicate that $pre(e) = \phi$, and p, $\bar{p}$ and $\not p$ on the second line of an event e to indicate that $post_e(p) = 1$, $post_e(p) = 0$ and $post_e(p) = \bot$, respectively.

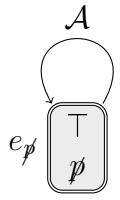

Figure 6.8: The event model $\mathcal{E}_{-p}$ for forgetting awareness.

Indeed, when applying the event model for forgetting awareness to an epistemic model $\mathcal{M}$, all worlds are preserved because, trivially, for each world $w \in \mathcal{M}$: $\mathcal{M}, w \models \top$ and hence $\mathcal{M}, w \not\models \top$. However, the valuation of p is deleted because $post_{e_{\not p}}(p) = \bot$. This corresponds to the model transformer $-p$ discussed in the previous section. Therefore, we will generally use $-p$ when referring to forgetting awareness, i.e. to denote the multipointed event model $\langle \mathcal{E}_{-p}, \{e_{\not p}\}\rangle$, both in the case of ParDEL$\odot$ and ParEAL.

Private forgetting awareness

With the event model for forgetting awareness, we can also define the event model for *private forgetting awareness*. Just like private announcements (Section 2.2.2 on page 35) and private raising awareness (Section 5.8.1 on page 106), private forgetting awareness for a group $G \subseteq \mathcal{A}$ causes the agents belonging to G to forget awareness, while it preserves the state of other agents not belonging to G.

However, different from the other private upgrades, we need to add two extra events to the event model for forgetting awareness in order to account for privacy. This is because we want this upgrade to preserve the properties of ParEAL, in particular specification. As such, it cannot be that from an event where the postcondition for p is assigned $\bot$, and hence its valuation is deleted, another event can be accessed where the postcondition for p is not defined, and therefore remains defined. Therefore, the event model for private forgetting awareness consists of three events representing these changes.

Definition 6.6 (Event Model for Private Forgetting Awareness)**.** The multipointed event model for *private forgetting awareness* of a proposition p is $\langle \mathcal{E}_{-Gp}, \{e^*\}\rangle$ where $\mathcal{E}_{-Gp} = \langle E_{-Gp}, (R_a)_{a\in\mathcal{A}}, pre, post\rangle$, with $E_{-Gp} = \{e^*, e_\top, e_{\not p}\}$, $R_a = \{\langle e^*, e_{\not p}\rangle, \langle e_{\not p}, e_{\not p}\rangle, \langle e_\top, e_\top\rangle\}$ for agents $a \in G$ and for agents $a \notin G$ $R_a = \{\langle e^*, e_\top\rangle, \langle e_{\not p}, e_{\not p}\rangle, \langle e_\top, e_\top\rangle\}$, preconditions $pre(e) = \top$ for every $e \in E_{-Gp}$ and postcondition $post_{e_{\not p}}(p) = \bot$ (see Figure 6.9).

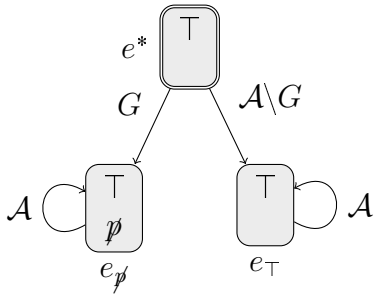

Figure 6.9: The event model $\mathcal{E}_{-_G p}$ for private forgetting awareness $-_G p$ for a group G of agents.

As before, we also use $E^*_{-_G p}$ to denote the point of reference $\{e^*\}$. We will also use $-_G p$ when referring to private forgetting awareness, and write $\mathcal{M}^{-_G p}$ for the product update $\mathcal{M} \otimes \mathcal{E}_{-_G p}$.

Private forgetting awareness could be used to forget awareness of a single agent when $G = \{a\}$, i.e. $-_{\{a\}} p$. In that case, only agent a has her awareness of p raised, and the other agents remain in the old state.

Because private forgetting awareness occurs, indeed, privately, other agents may mistakingly know that another agent knows or believes a proposition, even if that agent herself forgot awareness of it, see Figure 6.10. This is because for agents not belonging to the group G whose awareness is forgotten, the event is not observed. Hence, agents do not truly know what other agents know or believe.

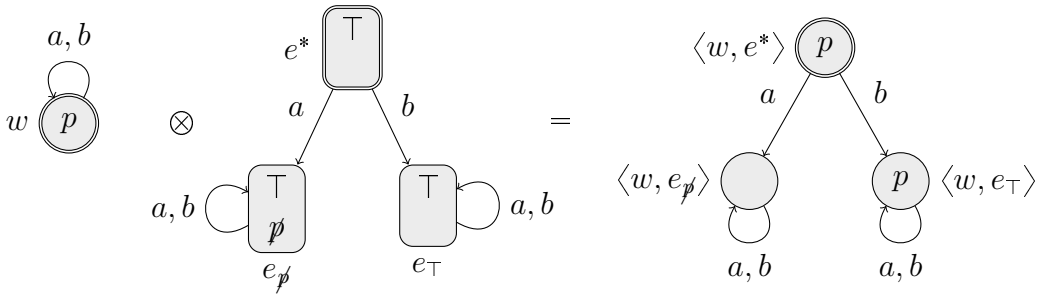

Figure 6.10: The event model of private forgetting awareness, $\mathcal{E}_{-_{\{a\}} p}$, applied to the epistemic model on the left. For simplicity, worlds in the product update that cannot be reached from $\langle w, e^* \rangle$ are omitted. Even though agent a has forgotten awareness of p, agent b still knows that a knows p. In other words, agent b does not know agent a lost awareness of p. Furthermore, agent a does not consider that agent b is aware of p.

Forgetting truth

Now, let us consider the forgetting truth modality. Defining an event model for forgetting truth is tricky because it is necessary to identify the worlds where p is undefined (and do nothing to these worlds), and the worlds where p is defined (and delete p there) via preconditions that are not falsified at these world, but are at other worlds.

For the worlds where p is undefined, such a precondition can be found quite easily: $p \wedge \neg p$. We can also refer to this formula as $\bot_p$ as it is a contradiction involving the proposition p. It then holds that $\mathcal{M}, w \not\models \bot_p$ if and only if $\mathcal{M}, w \not\models p \wedge \neg p$, if and only if $\mathcal{M}, w \models p$ and $\mathcal{M}, w \models \neg p$, or $\mathcal{M}, w \not\models p$ and $\mathcal{M}, w \not\models p$ – in which case p is undefined. As the first case cannot occur, it therefore holds that an event e with precondition $pre_e = \bot_p$ can only be applied to worlds where p is undefined. Thus, letting the precondition be $\bot_p$, only worlds where p is undefined are preserved.

Likewise, it would be a good guess to let the precondition for worlds where p is defined to be $p \vee \neg p$, which we can also refer to as $\top_p$. However, this does not only preserve worlds where p is defined, but also where p is undefined. Indeed, for all models $\mathcal{M}$ and worlds $w \in \mathcal{M}$: $\mathcal{M}, w \not\models \top_p$ if and only if $\mathcal{M}, w \not\models \neg\bot_p$ (because $\top_p = \neg\bot_p$), if and only if $\mathcal{M}, w \not\models \bot_p$, if and only if either $\mathcal{M}, w \not\models p$ or $\mathcal{M}, w \not\models p$ – thus in which cases p is true, false or undefined. Hence the event e with precondition $pre_e = \top_p$ preserves all worlds of the model.

Hence, we need to change the definition of the product update (Definition 5.16 on page 104). One way would be to use $\models$ instead of $\not\models$ in the clause where preconditions are used:

$$W^{\mathcal{M}\otimes\mathcal{E}} = \{\langle w, e\rangle \in W \times E \mid \mathcal{M}, w \models pre_e\}$$

So that an event e can be applied to a world w if w verifies the precondition of e. In this way, the precondition to $p \vee \neg p$ would select all the worlds in which p is defined (because $p \vee \neg p = \neg(p \wedge \neg p)$, which is verified only if p is true or false). However, it also causes the worlds where p is undefined to no longer be identified by $p \wedge \neg p$: this sentence is a contradiction so it can never be verified, it can only be *not falsified*.

Therefore, another approach is necessary. We can consider preconditions just like postconditions as partial functions $pre : E \to (P \to \{0, 1, \bot\})$, assigning to each event e and proposition p either 1, 0 or $\bot$, or let it be undefined. Then, the clause for $W^{\mathcal{M}\otimes\mathcal{E}}$ could be replaced as follows:

$$W^{\mathcal{M}\otimes\mathcal{E}} = \left\{ \langle w, e\rangle \in W \times E \;\middle|\; V_w(p) = \begin{cases} pre_e(p) & \text{if } pre_e(p) \in \{0,1\} \\ \text{undefined} & \text{if } pre_e(p) = \bot \\ V_w(p) & \text{otherwise} \end{cases} \right\}$$

So that a precondition assigning to p a value 1 would select worlds where p is true, a value 0 worlds where p is false and finally a value $\bot$ worlds where p is undefined. Furthermore, if for some proposition p the precondition is undefined for p, any world is preserved.

However, this complicates the event models for raising awareness and forgetting awareness we have seen thus far because it requires more events to be added: no longer a single event suffices to select the worlds where a proposition is defined (hence true or false).

Fortunately, there is another way. In the following section, we show how we can capture the forgetting truth modality with the modalities for private raising awareness and forgetting awareness, so that the event model for forgetting truth can be obtained by combining the event models of the other modalities.

6.3 A relation between the two types

In both types of forgetting, agents that are unaware of a proposition p are not affected by it. However, the functioning of the two types is quite different: $-p$ preserves all the worlds and only changes the valuation functions, whereas $\ominus p$ changes the structure of the model by duplicating all the worlds in which p in defined. In fact, this looks a lot like that of raising awareness: some worlds are duplicated, and the valuation function in these duplicated worlds is changed compared to the original world. In the case of $+p$, the worlds that are duplicated are those where p is undefined and the valuation function is extended with p, and in the case of $\ominus p$, the worlds that are duplicated are those where p is defined and the valuation function is flipped.

Here, we prove a relation between the three modalities: raising awareness, forgetting awareness and forgetting truth. We show that to perform $\ominus p$, we can first apply $-p$ followed by a private raising awareness modality $+_G p$ for a specific group G. The latter is private because there may be agents who are unaware of p that are not affected by $\ominus p$. Consider the example in Figure 6.11 in which agent a knows p and $\neg q$ and agent b only knows q (and is not aware of p). Forgetting truth of p causes agent a to become uncertain about p, whereas unawareness of agent b is preserved. On the other hand,

forgetting awareness of p and then raising awareness of it causes both agents to become uncertain about p and hence, also aware. As such, the raising awareness needs to occur only for those agents who were previously aware of p, hence $+_G p$ where G is the group of these agents. In Figure 6.11, this would correspond to $-p; +_{\{a\}} p$.

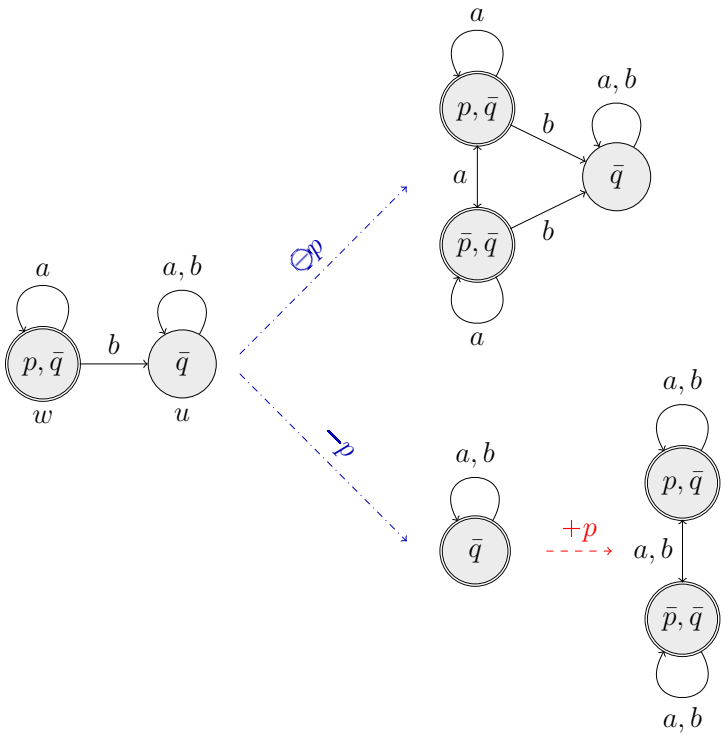

Figure 6.11: Forgetting truth, $\ominus p$, (above) and forgetting awareness followed by raising awareness, $-p; +p$, (below) applied to the model on the left. In the first case, agent a becomes uncertain about p, while agent b remains unaware of p, and in the second case, both agents become uncertain (and thus aware) about p. Note that in the model resulting from applying $-p; +p$, worlds have been merged up to bisimilarity.

Proposition 6.1. For any pointed ParDEL$\odot$ model $\langle \mathcal{M}, w \rangle$ for a set $\mathcal{A}$ of agents and any proposition $p \in P$, if $G \subseteq \mathcal{A}$ is the group of agents that are aware of p at w (and the agents $\mathcal{A} \backslash G$ are unaware of p), then $\mathcal{M}^{\ominus p}$ is bisimilar to $\mathcal{M}^{-p;+_G p}$.

Proof. Let $\langle \mathcal{M}, w \rangle$ be a pointed ParDEL$\odot$ model for a set $\mathcal{A}$ of agents and let p be a proposition. Then, let $G \subseteq \mathcal{A}$ be the group of agents who are aware of p. That is, $a \in G$ if and only if for all $v \in ||w||_a$: $p \in Dom(V_v)$.

Let Z be a relation between $\mathcal{M}^{\ominus p}$ and $\mathcal{M}^{-p;+_{GP}p}$ defined as: if $p \in Dom(V_v)$, then $\langle v, 1\rangle Z \langle v, e_p\rangle$ and $\langle v, 0\rangle Z \langle v, e_{\bar{p}}\rangle$, and if $p \notin Dom(V_v)$ then $\langle v, 0\rangle Z \langle v, e_\top\rangle$.

We show that Z is a bisimulation (Definition 2.24 on page 34):

- [**Propositional agreement**] For all $\langle w, i\rangle \in \mathcal{M}^{\ominus p}$ and $\langle w, e\rangle \in \mathcal{M}^{-p;+_{GP}p}$ such that $\langle w, i\rangle Z \langle w, e\rangle$:

$$
\begin{aligned}
V^{\ominus p}_{\langle w,i\rangle}(p) &= \begin{cases} 1 & i = 1, p \in Dom(V_w) \\ 0 & i = 0, p \in Dom(V_w) \\ \text{undefined} & p \notin Dom(V_w) \end{cases} \\
&= \begin{cases} V^{-p;+_{GP}p}_{\langle w,e_p\rangle}(p) & i = 1, p \in Dom(V_w) \\ V^{-p;+_{GP}p}_{\langle w,e_{\bar{p}}\rangle}(p) & i = 0, p \in Dom(V_w) \\ V^{-p;+_{GP}p}_{\langle w,e_\top\rangle}(p) & p \notin Dom(V_w) \end{cases} \\
&= V^{-p;+_{GP}p}_{\langle w,e\rangle}
\end{aligned}
$$

- [**Forth**] Let $\langle v, i\rangle \in \mathcal{M}^{\ominus p}$ and $\langle v, e\rangle \in \mathcal{M}^{-p;+_{GP}p}$ such that $\langle v, i\rangle Z \langle v, e\rangle$. Suppose that $\langle v, i\rangle R^{\ominus p}_a \langle v', j\rangle$ for some agent $a \in \mathcal{A}$.

  Let $\langle v', e'\rangle$ be such that if $p \in Dom(V_{v'})$ and $j = 1$: $e' = e_p$, if $p \in Dom(V_{v'})$ and $j = 0$: $e' = e_{\bar{p}}$, and if $p \notin Dom(V_{v'})$: $e' = e_\top$. In each case $\langle v', e'\rangle \in \mathcal{M}^{-p;+_{GP}p}$ because after applying $-p$, p is undefined everywhere meaning that all three events e_p, $e_{\bar{p}}$ and $e_\top$ can be applied to v'. Furthermore, clearly, $\langle v', j\rangle Z \langle v', e'\rangle$. It remains to show that $\langle v, e\rangle R^{-p;+_{GP}p}_a \langle v', e'\rangle$, see Figure 6.12.

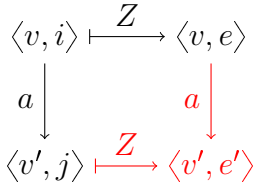

Figure 6.12: A visualization of the 'forth' condition of the bisimulation: given the worlds on the top and bottom left and the relations Z and R_a between them on the top and left, respectively (black), we need to show that the world on the bottom right exists with the relations Z and R_a on the bottom and right, respectively (red).

  By $\langle v, i\rangle R^{\ominus p}_a \langle v', j\rangle$, and since $\ominus p$ preserves relations, it must be that $v R_a v'$. Then, $\langle v, e\rangle R^{-p;+_{GP}p}_a \langle v', e'\rangle$ holds if and only if $e R^{\mathcal{E}_{+p}}_a e'$. There are two cases:

1. $p \in Dom(V_v)$: then $e \in \{e_p, e_{\bar{p}}\}$. There are now two options, $p \in Dom(V_{v'})$ (hence $e' \in \{e_p, e_{\bar{p}}\}$) or $p \notin Dom(V_{v'})$ (hence $e' = e_\top$). Furthermore, in the first case, it must be that $a \in G$, whereas in the second case it must be that $a \notin G$. In both cases, $eR_a^{\mathcal{E}+p}e'$.
2. $p \notin Dom(V_v)$: then, by specification (Definition 5.4 on page 92), it must be that $p \notin Dom(V_{v'})$ and therefore $e = e' = e_\top$. And $e_\top R_a^{\mathcal{E}+p} e_\top$ for all $a \in \mathcal{A}$.

- **[Back]** Let $\langle v, i\rangle \in \mathcal{M}^{\ominus p}$ and $\langle v, e\rangle \in \mathcal{M}^{-p;+Gp}$ such that $\langle v, i\rangle Z \langle v, e\rangle$. Suppose that $\langle v, e\rangle R_a^{-p;+Gp}\langle v', e'\rangle$ for some agent $a \in \mathcal{A}$.

  Let $\langle v', j\rangle$ be such that if $e' = e_p$ and $p \in Dom(V_{v'})$: $j = 1$, and otherwise $j = 0$. We need to show that indeed $\langle v', j\rangle \in \mathcal{M}^{\ominus p}$, $\langle v, i\rangle R_a^{\ominus p}\langle v', j\rangle$ and $\langle v', j\rangle Z \langle v', e'\rangle$, see Figure 6.13.

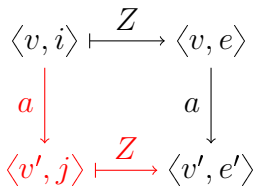

Figure 6.13: A visualization of the 'back' condition of the bisimulation: given the worlds on the top and bottom right and the relations Z and R_a between them on the top and right, respectively (black), we need to show that the world on the bottom left exists with the relations Z and R_a on the bottom and left, respectively (red).

  Suppose $\langle v', j\rangle \notin \mathcal{M}^{\ominus p}$. Since for each $u \in \mathcal{M}$, $\langle u, 0\rangle \in \mathcal{M}^{\ominus p}$ this only occurs if $j = 1$ and $p \notin Dom(V_{v'})$. But such a $\langle v', j\rangle$ does not occur by construction: either (1) $p \in Dom(V_{v'})$, $e = e_p$ and $j = 1$, or (2) $j = 0$. Hence it must be that $\langle v', j\rangle \in \mathcal{M}^{\ominus p}$.

  Then, because $\langle v, e\rangle R_a^{-p;+Gp}\langle v', e'\rangle$, it must be that vR_av'. And since $\ominus p$ preserves relations, also $\langle v, i\rangle R_a^{\ominus p}\langle v', j\rangle$.

  Finally, because of how we choose $\langle v', j\rangle$, it holds by definition of Z that $\langle v', j\rangle Z \langle v', e'\rangle$.

Hence $\mathcal{M}^{\ominus p}$ is bisimilar to $\mathcal{M}^{-p;+Gp}$. □

Proposition 6.1 also suggests that private forgetting truth $\ominus_G p$ for a group $G \subseteq \mathcal{A}$ can be captured by performing $-_G p$ followed by $+_{G'} p$, where $G' \subseteq G$

are the sub-group of agents in G already aware of p in the initial model. This is because then forgetting awareness will only be applied to the agents in G, hence agents not in G who were aware of p remain so, after which the agents in G who were already aware of p raise their awareness and thus become uncertain. This will be useful when we turn back to the model of the Alignment Repair Game.

6.4 Forgetting awareness implies forgetting truth

Forgetting awareness implies forgetting truth. This is for the obvious reason that the forgetting awareness modality deletes the valuation of propositions and hence, they cannot remain true or false.

Proposition 6.2 (Forgetting awareness implies forgetting truth)**.** Let $\mathcal{M}$ be a ParDEL$\odot$ model and let w be a state in $\mathcal{M}$. Then for any proposition $p \in P$:

$$\mathcal{M}, w \not\models [-p]p \text{ and } \mathcal{M}, w \not\models [-p]\neg p$$

Proof. Let $\mathcal{M}$ be a ParDEL$\odot$ model, w a state in $\mathcal{M}$ and p a proposition. Then $\mathcal{M}^{-p}$ is just like $\mathcal{M}$, but where the valuation of p is deleted. That is, $V_w^{-p}(p)$ is undefined. But then, $\mathcal{M}^{-p}, w \not\models p$ and $\mathcal{M}^{-p}, w \not\models p$, i.e. $\mathcal{M}^{-p}, w \not\models \neg p$. Hence, $\mathcal{M}, w \not\models [-p]p$ and $\mathcal{M}, w \not\models [-p]\neg p$. □

We can also use the forgetting truth modality $\ominus p$ to show that forgetting awareness implies forgetting truth.

Proposition 6.3. Let $\mathcal{M}$ be a ParDEL$\odot$ model. Then for any proposition $p \in P$: $\mathcal{M}^{-p;\ominus p}$ is bisimilar to $\mathcal{M}^{-p}$.

Proof. We can define Z such that $\forall w \in \mathcal{M}$: wZw. It is then obvious that Z is a bisimulation by observing that after removing the valuation of p via $-p$, there is no world w left such that $p \in Dom(V_w)$. Hence $\ominus p$ does not alter the model. □

For the forgetting truth modality, awareness is not lost, but knowledge and beliefs are. Agents become uncertain.

Proposition 6.4 (Forgetting truth makes uncertain)**.** Let $\mathcal{M}$ be a ParDEL$\odot$ model and let w be a state in $\mathcal{M}$. Then for any proposition $p \in P$:

$$\mathcal{M}, w \not\models [\ominus p]K_a p \text{ and } \mathcal{M}, w \not\models [\ominus p]K_a \neg p$$

$$\mathcal{M}, w \not\models [\ominus p]B_a p \text{ and } \mathcal{M}, w \not\models [\ominus p]B_a \neg p$$

Proof. Let $\mathcal{M}$ be a ParDEL$\odot$ model and p be a proposition. Then $\mathcal{M}^{\ominus p}$ is $\mathcal{M}$ with every world w where $p \in Dom(V_w)$ duplicated, such that p is true in one world, and false in the other. Otherwise, these worlds are equivalent, satisfying the same relations and same valuations of other propositions. But then, it can never be that an agent knows p or $\neg p$ because, whenever she has access to a p-world, she also has access to another $\neg p$-world. Likewise, agents cannot believe p or $\neg p$ because for any most plausible p-world, there is another, equally plausible, $\neg p$-world available to her. Hence, agents are uncertain about p in $\mathcal{M}^{\ominus p}$. □

6.5 Conclusion

In this chapter, two modalities for forgetting have been introduced. A forgetting awareness modality that reduces the scope of the valuation function by deleting the valuation of the proposition that is to be forgotten, and a forgetting truth modality that copies worlds in which the proposition that is to be forgotten is defined and the truth value is flipped. We have defined event models for the modalities and the relation between the two types has been explored.

The link between forgetting and ARG is that adaptive agents use objects to evolve their alignment, but do not remember the class of the object that was announced. That is, they forget this information after using it for a more general purpose. Experiments [46, 49] have shown that alignments are sufficient for agents to reach successful communication and that they do not need the objects for this purpose. Thus, incorporating forgetting into the translation of the adaptation operators could bring the logical model closer to ARG by allowing the logical agents to forget the class of the objects that were used to evolve the alignment, like in the original game. As a consequence, the formal properties of the adaptation operators need to be revisited with respect to the new translation. In particular, this may affect incompleteness because the proof of incompleteness builds on the ability of logical agents to remember the classifications of the objects. The next chapter will explore this by defining a new translation in which awareness is used.

Chapter 7

Formal Properties of the Adaptation Operators Revisited

> Most of the mistakes in thinking
> are inadequacies of perception
> rather than mistakes of logic.
>
> Edward de Bono

The previous chapters introduced awareness and defined raising awareness and forgetting modalities. In this chapter, we take these ideas back to the Alignment Repair Game (ARG), by providing an alternative logical model of ARG and re-evaluating the formal properties of the adaptation operators with respect to this model. This is achieved through defining Partial Dynamic Epistemic Ontology Logic (ParDEOL), which combines the notion of awareness and modalities for raising awareness and forgetting in ParDEL with the encoding of ontologies and alignments in DEOL. Together, they ensure that agents can use ontologies to represent their knowledge and that agents do not necessarily have full awareness of the terms used by other agents in their ontologies.

A new translation from ARG to ParDEOL is defined and the formal properties (correctness, redundancy and completeness) of the adaptation operators are re-examined with respect to this translation. It is shown that the adaptation operators are correct and complete with respect to this translation, and partial redundancy does no longer hold, showing that indeed the logical model of ARG in DEOL can be considered insufficient to capture the behavior of agents and an alternative logic is required.

Furthermore, with a formal notion of awareness, it is explored how awareness evolves through playing ARG and it is shown that agents become aware only of the sub-vocabularies necessary to succeed in the game, which are not necessarily the full vocabularies.

7.1 Partial Dynamic Epistemic Ontology Logic

Partial Dynamic Epistemic Ontology Logic (ParDEOL) is the extension of Dynamic Epistemic Logic (DEL) using propositions of a simple Description Logic language (see Section 3.1) and defined with respect to weakly reflexive relations and interpretations like Partial Dynamic Epistemic Logic (ParDEL) (see Chapter 5).

Its syntax is defined with announcements, conservative upgrades and awareness upgrades for classes (raising awareness and forgetting awareness). It can be extended with radical upgrades, but as they will not be needed in the translation of ARG to ParDEOL, we leave them out for simplicity.

Definition 7.1 (Syntax of ParDEOL)**.** Given a countable, non-empty set $\mathcal{C}$ of class names, a countable, non-empty set $\mathcal{D}$ of object names, and a finite, non-empty set $\mathcal{A}$ of agents, the *syntax*, $\mathcal{L}_{ParDEOL}$, of (multi-agent) Partial Dynamic Epistemic Ontology Logic (ParDEL) is defined in the following way.

$$\phi ::= C(o) \mid CRD \mid \phi \wedge \psi \mid \neg\phi \mid K_a\phi \mid B_a\phi \mid [!\phi]\phi \mid \mid [\uparrow\phi]\phi \mid \mid [\odot_G C]\phi$$

$$R \in \{\sqsubseteq, \sqsupseteq, \oplus\}, \odot \in \{+, -, \ominus\}$$

where $C, D, \top \in \mathcal{C}$, $o \in \mathcal{D}$, K_a and B_a are the knowledge and belief operators for agent a and $G \subseteq \mathcal{A}$.

The connectives $\rightarrow$ and $\vee$ are defined as usual.

ParDEOL frames are DEOL frames (Definition 3.2 op page 48) with weakly reflexive relations instead of reflexive relations. Therefore the relations are called *accessibility relations* rather than plausibility relations and are denoted by R_a.

Definition 7.2 (ParDEOL Frames)**.** Given a finite, non-empty set $\mathcal{A}$ of agents, a *frame* of (multi-agent) ParDEOL is a pair $\mathfrak{F} = \langle W, (R_a)_{a\in\mathcal{A}}\rangle$ where

- W is a non-empty set of worlds, and
- $(R_a)_{a\in\mathcal{A}} \subseteq W \times W$ are the accessibility relations on W, one for each agent, that are well-founded, locally connected, weakly reflexive and transitive.

From the accessibility relations, the epistemic and doxastic relations can be defined in the same way as for ParDEL with respect to the aware cell of an agent (Definition 5.6 on page 5.6):

$$w \sim_a v \quad \text{iff} \quad v \in ||w||_a \tag{7.1}$$

$$w \rightarrow_a v \quad \text{iff} \quad v \in Max_{R_a}||w||_a \tag{7.2}$$

ParDEOL models are ParDEOL frames equipped with a partial interpretation function satisfying *consideration consistency* and *specification*.

Definition 7.3 (Consideration Consistency and Specification for Interpretations)**.** Let W be a non-empty set of worlds, $\mathcal{C}$ be a countable, non-empty set of class names, Δ a domain and $R_a \subseteq W \times W$ an accessibility relation on W. Then for any interpretation I that assigns to each world $w \in W$ a partial function $\cdot^{I_w} : \mathcal{C} \rightarrow \mathcal{P}(\Delta)$:

I satisfies *consideration consistency* if $\forall w, v, u \in W$:

$$wR_av \wedge wR_au \Rightarrow Dom(I_v) = Dom(I_u) \tag{7.3}$$

I satisfies *specification* if $\forall w, v \in W$:

$$wR_av \Rightarrow Dom(I_v) \subseteq Dom(I_w) \tag{7.4}$$

Definition 7.4 (ParDEOL Model)**.** Given a countable, non-empty set $\mathcal{C}$ of class names, a countable, non-empty set $\mathcal{D}$ of object names, and a finite, non-empty set $\mathcal{A}$ of agents, a *model* of (multi-agent) ParDEOL is a triple $\mathcal{M} = \langle \mathfrak{F}, \Delta, I \rangle$ where

- $\mathfrak{F}$ is a ParDEOL frame,
- Δ is the domain of interpretation, and
- I is an *interpretation function* assigning to each world $w \in W$ a partial function $\cdot^{I_w}$ such that $\cdot^{I_w}$ assigns to object names $o \in \mathcal{D}$ an element of the domain Δ ($\cdot^I : \mathcal{D} \rightarrow \Delta$), and to class names $C \in \mathcal{C}$ a subset of Δ ($\cdot^I : \mathcal{C} \rightarrow \mathcal{P}(\Delta)$), satisfying consideration consistency, specification and $\top^{I_w} = \Delta$.

7.1.1 Dynamic Upgrades for ParDEOL

Announcements and conservative upgrades are defined as for ParDEL (Section 5.7 on page 101): deleting all $\neg\phi$-worlds and pushing the ϕ-worlds on top of $\neg\phi$-worlds in the aware cells of agents, respectively. Here, we define the awareness upgrades with respect to event models.

Like for DEL and ParDEL, event models for ParDEOL are relational structures for dynamic modalities, in the same way as (Kripke) models are for static information. The difference is that they are defined for classes instead of propositions. This requires to define postconditions as partial functions $post : E \to (\mathcal{C} \to (\mathcal{P}(\Delta) \cup \{\bot\}))$ such that for each event $e \in E$ and class $C \in \mathcal{C}$, either a subset of the domain Δ is assigned, $\bot$ is assigned, or it is undefined. Subsets of Δ will be assigned to extend the valuation of a class C, so that awareness of it can be raised, $\bot$ will be assigned to delete the valuation of a class C, so that classes can be forgotten, and it will be undefined to preserve the old interpretation of C, so that nothing happens.

Definition 7.5 (Event Model for ParDEOL)**.** Let $\mathcal{C}$ be a countable, non-empty set of class names, a countable, Δ be a countable, non-empty set of object names, and $\mathcal{A}$ be a finite, non-empty set of agents. An *event model* for DEL is a quadruple $\mathcal{E} = \langle E, (R_a)_{a\in\mathcal{A}}, pre, post\rangle$ where

- E is a non-empty set of *events*,
- $(R_a)_{a\in\mathcal{A}} \subseteq E \times E$ are the *accessibility relations* on E, one for each agent $a \in \mathcal{A}$,
- $pre : E \to \mathcal{L}_{ParDEOL}$ is a *precondition function* assigning to each event a formula ϕ, and
- $post : E \to \big(\mathcal{C} \to (\mathcal{P}(\Delta) \cup \{\bot\})\big)$ is a *postcondition function* assigning to each event a partial function $post_e : \mathcal{C} \to \{\varnothing, \ldots, \Delta, \bot\}$.

A *pointed event model (for ParDEOL)* is a pair $\langle \mathcal{E}, e\rangle$ where $\mathcal{E}$ is an event model for ParDEOL and $e \in E$.

A *multipointed event model (for ParDEOL)* is a pair $\langle \mathcal{E}, E^*\rangle$ where $\mathcal{E}$ is an event model for ParDEOL and $E^* \subseteq E$.

We will again also write pre_e for $pre(e)$ and $post_e(p)$ for $(post(e))(p)$.

Product updates for ParDEOL are like those for ParDEL, except that instead of valuations, interpretations are affected: whenever the postcondition for a class C is defined and not equal to $\bot$, this becomes the new valuation of C, whenever the postcondition is equal to $\bot$, the class C becomes undefined and whenever the postcondition is undefined, the old interpretation of C is preserved.

Definition 7.6 (Product Update for ParDEOL)**.** Let $\mathcal{M} = \langle W, (R_a^{\mathcal{M}})_{a\in\mathcal{A}}, \Delta, I\rangle$ be a ParDEOL model and $\mathcal{E} = \langle E, (R_a^{\mathcal{E}})_{a\in\mathcal{A}}, pre, post\rangle$ be an event model for ParDEOL. Their *product update* $\mathcal{M}\otimes\mathcal{E} = \langle W^{\mathcal{M}\otimes\mathcal{E}}, (R_a^{\mathcal{M}\otimes\mathcal{E}})_{a\in\mathcal{A}}, \Delta^{\mathcal{M}\otimes\mathcal{E}}, I^{\mathcal{M}\otimes\mathcal{E}}\rangle$ is defined by:

- $W^{\mathcal{M}\otimes\mathcal{E}} = \{\langle w, e\rangle \in W \times E \mid \mathcal{M}, w \not\models pre(e)\}$
- $\langle w, e\rangle R_a^{\mathcal{M}\otimes\mathcal{E}}\langle w', e'\rangle$ iff $\langle w, e\rangle, \langle w', e'\rangle \in W^{\mathcal{M}\otimes\mathcal{E}}$, $wR_a^{\mathcal{M}}w'$ and $eR_a^{\mathcal{E}}e'$
- $\Delta^{\mathcal{M}\otimes\mathcal{E}} = \Delta$
- $I^{\mathcal{M}\otimes\mathcal{E}}_{\langle w,e\rangle}(C) = \begin{cases} post_e(C) & \text{if } post_e(C) \in \{\varnothing, \ldots, 2^{|\Delta|}\} \\ \text{undefined} & \text{if } post_e(C) = \bot \\ I_w(C) & \text{otherwise} \end{cases}$

Raising class awareness

Raising (public) awareness of a class C, denoted by $+_{\mathcal{A}}C$ (or short: $+C$), is achieved by creating $2^{|\Delta|}$ copies of the worlds in which C is undefined (i.e. $C \notin Dom(I_w)$ for world w), assigning to each these copies a different interpretation of C as subsets of Δ: from $\varnothing, \{o_1\}, \{o_2\}, \{o_1, o_2\}, \ldots$ up until the whole domain Δ. Other than the different interpretation of C (and thus related sentences such as $C(o)$, CRD, $K_a(C(o)$, etc.), these copied worlds are indifferent: satisfying the same relations and same interpretations. Furthermore, worlds in which C is defined are preserved.

We also use $\mathcal{M}^{+C} := \mathcal{M} \otimes \mathcal{E}_{+C}$ to denote the model obtained from applying the event model $\mathcal{E}_{+C}$ to $\mathcal{M}$, and W^{+C} to denote $W^{\mathcal{M}\otimes\mathcal{E}_{+C}}$.

Definition 7.7 (Event Model for Raising Class Awareness)**.** The multi-pointed event model for *raising class awareness* of a class C is the pair $\langle\mathcal{E}_{+C}, \{e_1, \ldots, e_{2|\Delta|}\}\rangle$ where $\mathcal{E}_{+C} = \langle E_{+C}, (R_a)_{a\in\mathcal{A}}, pre, post\rangle$, with $E_{+C} = \{e_1, \ldots, e_{2|\Delta|}\}$, $R_a = E_{+C} \times E_{+C}$ for $a \in \mathcal{A}$, and the pre- and postconditions are defined as follows (see Figure 7.1):

- $pre(e_1) = \bigwedge_{o\in\Delta}(\neg C(o))$, $post_{e_1}(C) = \varnothing$
- $pre(e_2) = C(o_1) \wedge \bigwedge_{o\in\Delta\setminus\{o_1\}}(\neg C(o))$, $post_{e_2}(C) = \{o_1\}$
- $\ldots$
- $pre(e_{2|\mathcal{D}|}) = \bigwedge_{o\in\Delta}(C(o))$, $post_{e_{2|\Delta|}}(C) = \Delta$

As before, we also use E^*_{+C} to denote the points of reference $\{e_1, \ldots, e_{2|\Delta|}\}$.

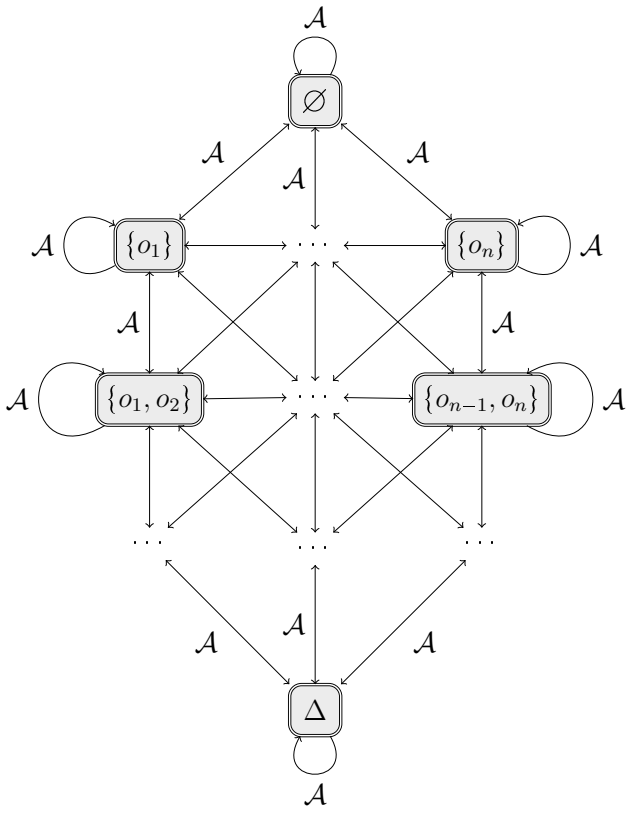

Figure 7.1: The event model $\mathcal{E}_{+C}$ for raising class awareness $+C$ for a domain $\Delta = \{o_1, \ldots, o_n\}$. Because the pre- and postconditions 'coincide' (that is, the valuation defined by the postcondition verifies the precondition), $\Delta' \subseteq \Delta$ is written inside an event e to denote $pre(e) = \bigwedge_{o \in \Delta'}(C(o)) \wedge \bigwedge_{o \in \Delta \setminus \Delta'}(\neg C(o))$ and $post_e(C) = \Delta'$.

Raising private class awareness of a class C, denoted by $+_G C$ where $G \subseteq \mathcal{A}$, is like private raising awareness for ParEAL (Definition 5.19 on page 107), by adding an event $e_\top$ to the event model for raising class awareness with precondition $\top$ and no postcondition. For all agents $a \in G$, the same relations hold as for $\mathcal{E}_{+C}$, and for all agents $a \notin G$, the relations are $(E_{+C} \cup \{e_\top\}) \times \{e_\top\}$. This is to ensure that these agents do not observe the event of raising awareness and still consider the old situation.

We also use $\mathcal{M}^{+_G C} := \mathcal{M} \otimes \mathcal{E}_{+_G C}$ to denote the model obtained from applying the event model $\mathcal{E}_{+_G C}$ to the ParDEOL model $\mathcal{M}$, and $W^{+_G C}$ to denote $W^{\mathcal{M} \otimes \mathcal{E}_{+_G C}}$.

Definition 7.8 (Event Model for Private Raising Class Awareness)**.** The multipointed event model for *raising private class awareness* of a class C for a group $G \subseteq \mathcal{A}$ is $\langle \mathcal{E}_{+_G C}, \{e_1, \ldots, e_{2^{|\Delta|}}\}\rangle$ where $\mathcal{E}_{+_G C} = \langle E_{+_G C}, (R_a)_{a\in\mathcal{A}}, pre, post\rangle$, with $E_{+_G C} = \{e_1, \ldots, e_{2^{|\Delta|}}, e_\top\}$, $R_a = \{e_1, \ldots, e_{2^{|\Delta|}}\}\times\{e_1, \ldots, e_{2^{|\Delta|}}\}\cup\{\langle e_\top, e_\top\}$ for $a \in G$ and $R_a = E_{+_G C} \times \{e_\top\}$ for $a \notin G$, and the pre- and postconditions are defined as follows:

- $pre(e_1) = \bigwedge_{o\in\Delta}(\neg C(o))$, $post_{e_1}(C) = \varnothing$
- $pre(e_2) = C(o_1) \wedge \bigwedge_{o\in\Delta\setminus\{o_1\}}(\neg C(o))$, $post_{e_2}(C) = \{o_1\}$
- $\ldots$
- $pre(e_{2^{|\Delta|}}) = \bigwedge_{o\in\Delta}(C(o))$, $post_{e_{2^{|\Delta|}}}(C) = \Delta$
- $pre(e_\top) = \top$

As before, we also use $E^*_{+_G C}$ to denote the points of reference $\{e_1, \ldots, e_{2^{|\Delta|}}\}$.

Forgetting class awareness

Forgetting (public) awareness of a class C, denoted by $-_{\mathcal{A}}C$ (or short: $-C$), is achieved by reducing the scope of the interpretations by C. That is, the interpretation of C is deleted in all worlds of the model and C becomes undefined.

We also use $\mathcal{M}^{-C} := \mathcal{M} \otimes \mathcal{E}_{-C}$ to denote the model obtained from applying the event model $\mathcal{E}_{-C}$ to the ParDEOL model $\mathcal{M}$, and W^{-C} to denote $W^{\mathcal{M}\otimes\mathcal{E}_{-C}}$.

Definition 7.9 (Event Model for Forgetting Class Awareness)**.** The multipointed event model for *forgetting class awareness* of a class C is the pair $\langle \mathcal{E}_{-C}, e_\top\rangle$ where $\mathcal{E}_{-C} = \langle E_{-C}, (R_a)_{a\in\mathcal{A}}, pre, post\rangle$, with $E_{-C} = \{e_\top\}$, $R_a = \{\langle e_\top, e_\top\rangle\}$ for all $a \in \mathcal{A}$, $pre(e_\top) = \top$ and $post_{e_\top}(C) = \bot$ (see Figure 7.2).

As before, we also use E^*_{-C} to denote the point of reference $\{e_\top\}$.

When drawing event models, like before, we write ϕ on the first line of an event e to indicate that $pre(e) = \phi$, and now also write $\not C$ on the second line of an event e to indicate that $post_e(C) = \bot$.

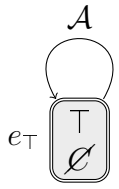

Figure 7.2: The event model $\mathcal{E}_{-C}$ for forgetting awareness $-C$.

Examples of applying the modality for forgetting class awareness look like those for forgetting awareness (Section 6.2.1 on page 122), with the difference that in this case, a lot more events are involved.

Private forgetting class awareness of a class C, denoted by $-_G C$, is like private forgetting awareness for ParEAL (Definition 6.6 on page 123), by adding two events to the event model for forgetting awareness with preconditions $\top$ and no postcondition. Two events are added instead of one in order to preserve the properties of ParDEOL models, in particular consideration consistency. One of these added event will be used for satisfiability, and the other to denote that agents not in G do not observe the forgetting.

We also use $\mathcal{M}^{-_G C} = \mathcal{M} \otimes \mathcal{E}_{-_G C}$ to denote the model obtained from applying the event model $\mathcal{E}_{-_G C}$ to the ParDEOL model $\mathcal{M}$, and $W^{-_G C}$ to denote $W^{\mathcal{M} \otimes \mathcal{E}_{-_G C}}$.

Definition 7.10 (Event Model for Private Forgetting Class Awareness)**.** The multipointed event model for *private forgetting class awareness* of a class C is the pair $\langle \mathcal{E}_{-_G C}, e^* \rangle$ where $\mathcal{E}_{-_G C} = \langle E_{-_G C}, (R_a)_{a \in \mathcal{A}}, pre, post \rangle$, with $E_{-_G C} = \{e^*, e_\top, e_{\not\subset}\}$, relations, for agents $a \in G$, $R_a = \{\langle e^*, e_{\not\subset}\rangle, \langle e_{\not\subset}, e_{\not\subset}\rangle, \langle e_\top, e_\top\rangle\}$ and, for agents $a \notin G$, $R_a = \{\langle e^*, e_\top\rangle, \langle e_{\not\subset}, e_{\not\subset}\rangle, \langle e_\top, e_\top\rangle\}$. preconditions $pre(e) = \top$ for every $e \in E_{-_G C}$ and postcondition $post_{e_{\not\subset}}(C) = \bot$ (see Figure 6.9).

As before, we also use $E^*_{-_G C}$ to denote the point of reference $\{e^*\}$.

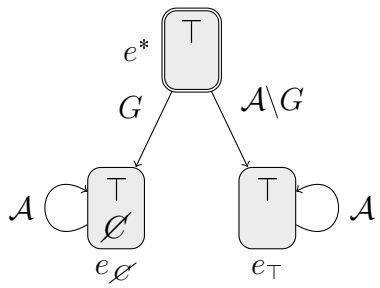

Figure 7.3: The event model $\mathcal{E}_{-_G C}$ for private forgetting class awareness $-_G C$ for a group G of agents.

Forgetting class truth

Like for ParDEL, we consider a second forgetting modality that we can call *forgetting class truth* $\ominus C$. This modality creates $2^{|\Delta|}$ copies all the worlds in which C is defined and assigns different subsets of Δ to these copied worlds. This means that whatever agents knew or believed about C will be dropped, while awareness is preserved – C remains interpreted.

For the propositional case, we discussed in Section 6.3 (page 126) how we can capture forgetting truth with raising and forgetting modalities. Therefore, we here define the forgetting class truth modality as an abbreviation: $\ominus C = -C; +_G C$ for a forgetting class awareness modality $-C$ followed by a private forgetting class awareness modality $-_G C$, where G are the group of agents previously aware of C. Likewise, we use $\ominus_G C$ to denote $-_G C; +_{G'} C$ where $G \subseteq G$ is the sub-group of agents in G previously aware of C. The abbreviation for private forgetting class truth will be used in the new translation of ARG.

7.1.2 Satisfiability for ParDEOL

Satisfiability for ParDEOL is based on satisfiability for DEOL (Definition 3.4 on page 49), but is defined with respect to verification ($\models$) and falsification ($=\!|$), like ParDEL (Definition 5.10 on page 96). As usual, it is considered with respect to a pointed model $\langle \mathcal{M}, w \rangle$ which associates a ParDEOL model $\mathcal{M}$ with a world $w \in W$.

Definition 7.11 (Satisfiability for ParDEOL)**.** Satisfiability for ParDEOL is defined as:

$$
\begin{array}{ll}
\mathcal{M}, w \vDash C(o) & \text{iff } C \in Dom(I_w) \text{ and } o^{I_w} \in C^{I_w} \\
\mathcal{M}, w \vDash C \sqsubseteq D & \text{iff } C, D \in Dom(I_w) \text{ and } C^{I_w} \subseteq D^{I_w} \\
\mathcal{M}, w \vDash C \equiv D & \text{iff } C, D \in Dom(I_w) \text{ and } C^{I_w} = D^{I_w} \\
\mathcal{M}, w \vDash C \oplus D & \text{iff } C, D \in Dom(I_w) \text{ and } C^{I_w} \cap D^{I_w} = \varnothing \\
\mathcal{M}, w \vDash \phi \wedge \psi & \text{iff } \mathcal{M}, w \vDash \phi \text{ and } \mathcal{M}, w \vDash \psi \\
\mathcal{M}, w \vDash \neg\phi & \text{iff } \mathcal{M}, w \dashv \phi \\
\mathcal{M}, w \vDash K_a\phi & \text{iff } \forall v \text{ s.t. } w \sim_a v : \mathcal{M}, v \vDash \phi \\
\mathcal{M}, w \vDash B_a\phi & \text{iff } \forall v \text{ s.t. } w \rightarrow_a v : \mathcal{M}, v \vDash \phi \\
\mathcal{M}, w \vDash [!\phi]\psi & \text{iff } \mathcal{M}^{!\phi}, w \vDash \psi \\
\mathcal{M}, w \vDash [\uparrow\phi]\psi & \text{iff } \mathcal{M}^{\uparrow\phi}, w \vDash \psi \\
\mathcal{M}, w \vDash [\odot_G C]\phi & \text{iff } \forall e \in E^*_{\odot_G C} : \mathcal{M}, w \nvDash pre(e) \text{ implies } \mathcal{M}^{\odot_G C}, \langle w, e_p \rangle \vDash \phi
\end{array}
$$

for verification ($\vDash$) and:

$$
\begin{array}{ll}
\mathcal{M}, w \dashv C(o) & \text{iff } C \in Dom(I_w) \text{ and } o^{I_w} \notin C^{I_w} \\
\mathcal{M}, w \dashv C \sqsubseteq D & \text{iff } C, D \in Dom(I_w) \text{ and } C^{I_w} \nsubseteq D^{I_w} \\
\mathcal{M}, w \dashv C \equiv D & \text{iff } C, D \in Dom(I_w) \text{ and } C^{I_w} \neq D^{I_w} \\
\mathcal{M}, w \dashv C \oplus D & \text{iff } C, D \in Dom(I_w) \text{ and } C^{I_w} \cap D^{I_w} \neq \varnothing \\
\mathcal{M}, w \dashv \phi \wedge \psi & \text{iff } \mathcal{M}, w \dashv \phi \text{ and } \mathcal{M}, w \vDash \psi, \\
 & \quad \text{or } \mathcal{M}, w \vDash \phi \text{ and } \mathcal{M}, w \dashv \psi, \\
 & \quad \text{or } \mathcal{M}, w \dashv \phi \text{ and } \mathcal{M}, w \dashv \psi \\
\mathcal{M}, w \dashv \neg\phi & \text{iff } \mathcal{M}, w \vDash \phi \\
\mathcal{M}, w \dashv K_a\phi & \text{iff } \exists v \text{ s.t. } w \sim_a v : \mathcal{M}, v \dashv \phi \\
\mathcal{M}, w \dashv B_a\phi & \text{iff } \exists v \text{ s.t. } w \rightarrow_a v : \mathcal{M}, v \dashv \phi \\
\mathcal{M}, w \dashv [!\phi]\psi & \text{iff } \mathcal{M}^{!\phi}, w \dashv \psi \\
\mathcal{M}, w \dashv [\uparrow\phi]\psi & \text{iff } \mathcal{M}^{\uparrow\phi}, w \dashv \psi \\
\mathcal{M}, w \dashv [\odot_G C]\phi & \text{iff } \exists e \in E^*_{\odot_G C} : \mathcal{M}, w \nvDash pre(e) \text{ and } \mathcal{M}^{\odot_G C}\langle w, e_p \rangle \dashv \phi
\end{array}
$$

for falsification ($\dashv$), where $\mathcal{M}^{\odot_G C} = \mathcal{M} \otimes \mathcal{E}_{\odot_G C}$ and $E^*_{\odot_G C}$ is the reference set of $\mathcal{E}_{\odot_G C}$.

As before, a set of formulas is *inconsistent* if there is no pointed model verifying it. We say that a formula ϕ is a consequence of a set of formulas Γ

(written $\Gamma \models \phi$) if every pointed model $\langle \mathcal{M}, w \rangle$ verifying all formulas of Γ, also verifies ϕ.

Whenever a class C does not belong to the domain of the interpretation function at a world w, i.e. $C \notin Dom(I_w)$, then $\mathcal{M}, w \not\models C(o)$ and $\mathcal{M}, w \not\models C(o)$ for each object $o \in \mathcal{D}$, and $\mathcal{M}, w \not\models CRC'$ and $\mathcal{M}, w \not\models CRC'$ for any $C' \in \mathcal{C}$ and any relation $R \in \{\sqsubseteq, \sqsupseteq, \oplus\}$, and C is said to be *undefined* at w.

Because raising class awareness is analogous to raising (propositional) awareness, the results about $+p$ can be extended to $+C$. Thus, for any pointed ParDEOL model $\langle \mathcal{M}, w \rangle$, any agent $a \in \mathcal{A}$ and any class $C \in \mathcal{C}$:

- [**Truth preservation and correspondence**] Raising class awareness $+C$ preserves truth and corresponds for formulas ϕ not containing C, i.e. $\mathcal{M}, w \models \phi$ implies $\mathcal{M}, w \models [+C]\phi$ and the latter implies the former when ϕ does not contain C (Proposition 5.8 on page 110 for ParEAL).

- [**Raise class awareness without truth**] If C was undefined, then after raising class awareness $+C$, truth of C is not disclosed, i.e. we have $\mathcal{M}, w \dashv [+C]C(o)$ and $\mathcal{M}, w \dashv [+C]\neg C(o)$ for any object $o \in \Delta$ (Proposition 5.9 on page 111 for ParEAL).

Similarly, because the mechanism for forgetting class awareness is similar to that of forgetting (propositional) awareness, the results about $-p$ can be extended to $-C$. Thus, for any pointed ParDEOL model $\langle \mathcal{M}, w \rangle$, any agent $a \in \mathcal{A}$ and any class $C \in \mathcal{C}$:

- [**Forgetting class awareness implies forgetting truth**] After forgetting class awareness $-C$, no classification $C(o)$ remains true or false, i.e. $\mathcal{M}, w \not\models [-C]C(o)$ and $\mathcal{M}, w \not\models [-C]\neg C(o)$ for any object $o \in \Delta$ (Proposition 6.2 on page 130 for ParEAL).

7.2 A New Translation

The reasons for adopting DEOL with awareness and the related raising and forgetting modalities was to bring the logical model of ARG closer to the original game. In ARG, adaptive agents use different vocabularies to express their knowledge, extend these vocabularies when encountering new terms and use the objects for the sole purpose of evolving their alignments, but are not interested in the class of these objects. This gave rise to ParDEOL, which combines the ideas of DEOL (using propositions from a Description Logic language to capture ontologies) and ParDEL (defining awareness with partial valuations/interpretations and weakly reflexive relations).

A different logic to model ARG requires a different translation from ARG to it. However, not completely: ARG states are translated to ParDEOL using the same translation τ as for DEOL (Definition 3.5 on page 51), mapping ontologies to agents' knowledge and alignments to agents' beliefs. The difference with DEOL arises when considering the semantics. In DEOL models of $\tau(s)$, each agent has full awareness of the vocabularies used by other agents, which is not true on ParDEOL models of $\tau(s)$. In fact, in ParDEOL models, each class not specified as knowledge or belief of an agent by the translation is a class the agent is unaware of. This means that for a pointed ParDEOL model $\langle \mathcal{M}, w \rangle$ of $\tau(s)$, any class C not appearing in the ontology or alignment of an agent is undefined in all the worlds accessible by her, i.e. $C \notin Dom(I_v)$ for any $v \in \mathcal{M}$ such that $w \sim_a v$ or $w \rightarrow_a v$. Hence, she does not know or believe anything about such C.

Awareness allows agents to use (be aware) of different vocabularies to represent their knowledge and beliefs. However, it also implies that we cannot translate the adaptation operators directly to announcements and conservative upgrades like for DEOL. This is because some of the classes to which a new correspondence is added are classes agents are unaware of. Therefore their awareness needs to be raised beforehand. This is not necessary for the class C_b that is announced, as it already appeared in the correspondence, but it is for the super- and subclasses of C_a and C_b to which correspondences are added.

Furthermore, different from the translation in DEOL, the logical translation of adaptation operators in ParDEOL makes use of a forgetting class truth modality after the announcement $!C_b(o)$. This corresponds to ARG, where agents use the announcement $!C_b(o)$ to test the correspondence and then discard it again in order to apply an adaptation operator to evolve the alignment. The forgetting class truth modality must occur before performing the conservative upgrade (forcing agents to believe the correspondence(s) added by the adaptation operator) and privately. It must occur before the conservative upgrade because, otherwise, what was learned through the conservative upgrade will likewise be forgotten, hence causing the new correspondence(s) not to be believed. Furthermore, it must occur privately because, otherwise, agent b will forget $C_b(o)$, which is part of her ontology, therefore violating the translation. In ParDEOL, the forgetting class truth modality is private to all the agents except b, because C_b occurs in agent b's ontology so she cannot forget: $\ominus_{\mathcal{A}\setminus\{b\}} C_b$.

The translation of adaptation operators in ParDEOL is denoted as δ^+.

Definition 7.12 (ARG Dynamics in ParDEOL)**.** Let $T = \tau(s)$ be the theory that is the translation of an ARG state s, let α be an adaptation operator and let $\langle C_a, C_b, \sqsupseteq\rangle$ be the failing correspondence for object o with $C_a \in \mathcal{O}_a$ and $C_b \in \mathcal{O}_b$. Then $\delta^+(\alpha[\langle C_a, C_b, \sqsupseteq\rangle, o]) : T \to T'$ is a theory transformation, where T' is defined as:

$$T' := \begin{cases} T \cup \{\langle !C_b(o)\rangle\top, \langle \ominus_{\mathcal{A}\setminus\{b\}} C_b\rangle\top\} & \text{if } \mathcal{O}_a \vDash C_a(o) \\ T \cup \{\langle !C_b(o)\rangle\top, \langle \ominus_{\mathcal{A}\setminus\{b\}} C_b\rangle\top, \big\langle d^+(\alpha[\langle C_a, C_b, \sqsupseteq\rangle, o])\big\rangle\top\} & \text{if } \mathcal{O}_a \nvDash C_a(o) \end{cases}$$

and $d^+(\alpha[\langle C_a, C_b, \sqsupseteq\rangle, o])$ is a complex logical upgrade that corresponds to the adaptation operator α applied to failing correspondence $\langle C_a, C_b, \sqsupseteq\rangle$ with object o, that is defined as $d(\alpha[\langle C_a, C_b, \sqsupseteq\rangle, o])$ (Definition 3.8 on page 61) proceeded by a raising class awareness modality for agent a:

$$\begin{aligned} d^+(\texttt{delete}) &= && d(\texttt{delete}) \\ d^+(\texttt{add}) &= +\, msc_a(C_a); && d(\texttt{add}) \\ d^+(\texttt{addjoin}) &= +\, mscc_a(o, C_a); && d(\texttt{addjoin}) \\ d^+(\texttt{refine}) &= \{+C'_b\}_{C'_b \in M_b(C_b,o)}; && d(\texttt{refine}) \\ d^+(\texttt{refadd}) &= +\, mscc_a(o, C_a); && \{+C'_b\}_{C'_b \in M_b(C_b,o)}; \; d(\texttt{refadd}) \end{aligned}$$

where $M_b(C_b, o) = \{C'_b \in mgcx_b(C_b, o) \mid \nexists\langle C'_a, C'_b, \sqsupseteq\rangle \in A_{ab}\}$.

7.3 Formal Properties of the Adaptation Operators Revisited

ParDEOL is introduced to overcome some differences between the adaptive agents playing ARG and the agents in the logical model of ARG in DEOL, as discussed in Chapter 4. In particular, (1) the notion of awareness in ParDEOL allows agents to use heterogeneous knowledge representations, which are based on different vocabularies, and (2) forgetting modalities enable agents to focus on general knowledge and discard the objects. To confirm that the ParDEOL model is closer to the original game, the formal properties of the adaptation operators need to be re-examined with respect to the new translation δ^+. Through this, the logical models of ARG in DEOL and in ParDEOL can be compared.

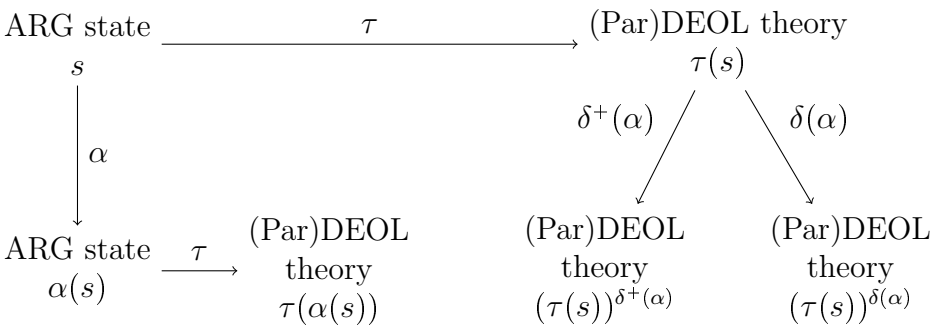

Figure 7.4: Relations between ARG scenarios and (Par)DEOL theories via the translation τ, adaptation operators α, DEOL dynamics $\delta(\alpha)$ and ParDEOL dynamics $\delta^+(\alpha)$.

7.3.1 Correctness

First, let us consider correctness. An adaptation operator α is correct with respect to a translation t if and only if for all ARG states s: $(\tau(s))^{t(\alpha)} \models \tau(\alpha(s))$ (Definition 3.9 on page 65). I.e. when applying the logical dynamics of the adaptation operator to the translated ARG state we arrive at a theory that entails the theory obtained from first applying the adaptation operator to the state and then translating it to (Par)DEOL. For the DEOL translation ($t = \delta$) it was proved that all adaptation operators but `add` are correct. This was because `add` does not take into account to which class the object o belongs and therefore the correspondence that is added to the alignment, with the direct superclass of C_a, might not be compatible with it, when this direct superclass is not a class of o (Section 3.3.1 on page 64).

With respect to the translation in ParDEOL ($t = \delta^+$), the situation has changed. This is because through the additional forgetting class awareness modality $\ominus C_b$ in the translation δ^+, agent a forgets that the object o belonged to C_b, which was the necessary ingredient for proving that `add` was incorrect for DEOL. Without this information, agent a can no longer conclude that the direct superclass of C_a, $msc_a(C_a)$, cannot subsume C_b in the case this superclass is not compatible with o: she still knows that o does not belong to $msc_a(C_a)$, but she no longer knows that o belongs to C_b. Hence, the conclusion is that all adaptation operators are correct with respect to the ParDEOL translation.

Proposition 7.1 (Correctness)**.** All adaptation operators are correct with respect to the ParDEOL translation δ^+.

Proof. To prove that the adaptation operators `delete`, `addjoin`, `refine` and `refadd` are correct, observe that through $\delta^+(\alpha)$, the same beliefs are added as for $\delta(\alpha)$, which contain the beliefs of the correspondence(s) added through α applied to s. Therefore, for the same reason that $\tau(s)^{\delta(\alpha)} \models \tau(\alpha(s))$ also $\tau(s)^{\delta^+(\alpha)} \models \tau(\alpha(s))$ for $\alpha \in \{\texttt{delete}, \texttt{addjoin}, \texttt{refine}, \texttt{refadd}\}$.

To prove that `add` is correct, we first look at the differences with the proof of incorrectness of `add` for δ (Proposition 3.5 on page 66). In this proof, it was shown that $\tau(s)^{!C_b(o);\Uparrow(C_a \sqsupset\!\sqsubset C_b \wedge msc_a(C_a) \sqsupseteq C_b)} \models K_a(C' \sqsupset\!\sqsubset C_b)$ but $\tau(\texttt{add}(s)) \models B_a(C' \sqsupseteq C_b)$ where C' is the direct superclass of C_a in agent a's ontology, i.e. $C' = msc_a(C_a)$. But this was based on the fact that the announcement $!C_b(o)$ causes agent a to know $C_b(o)$. However, when agent a forgets the extension of C_b before performing the conservative upgrade, this no longer holds, i.e. $\tau(s)^{!C_b(o);\ominus_{\mathcal{A}\setminus\{b\}}C_b;\Uparrow(C_a \sqsupset\!\sqsubset C_b \wedge msc_a(C_a) \sqsupseteq C_b)} \not\models K_a(C' \sqsupset\!\sqsubset C_b)$. Therefore, the argumentation in Proposition 3.5 (page 66) no longer holds. To show that add in incorrect, we then observe that the beliefs added by `add` are entailed: $\tau(s)^{!C_b(o);\ominus_{\mathcal{A}\setminus\{b\}}C_b;\Uparrow(C_a \sqsupset\!\sqsubset C_b \wedge C' \sqsupseteq C_b)} \models B_i(C_a \sqsupset\!\sqsubset C_b) \wedge B_i(C' \sqsupseteq C_b)$ for both agents $i \in \{a, b\}$ because of the conservative upgrade $\uparrow(C_a \sqsupset\!\sqsubset C_b \wedge msc_a(C_a) \sqsupseteq C_b)$. Since there are no other changes, it follows that `add` is correct. □

The correctness of `add` in the ParDEOL models of the translation compared to the incorrectness of `add` in the DEOL models of the translation leads to two observations: (1) the ParDEOL models are indeed closer to the original ARG game, but (2) the adaptive agents do benefit from more logical adaptation operators. This is because, even in ParDEOL, `add` remains a risky adaptation operator that may result in adding a faulty correspondence which has to be revised another time in the future. This occurs in exactly the case that led to incorrectness of `add` in DEOL: whenever the direct superclass $msc_a(C_a)$ of C_a, which belonged to the initial correspondence, is not compatible with o. Then, the correspondence added by `add` is $msc_a(C_a) \sqsupseteq C_b$, leading to another failure when o is drawn in a subsequent round. This goes on until the lowest superclass of C_a is reached that is compatible with o. This confirms the initial interpretation for the lack of formal properties with respect to the DEOL translation: either the adaptive agents use sub-logical behavior, or the logical model is insufficient to describe them. Turning to an alternative logic brings the model closer to the game, while dropping `add` brings the game closer to the logic.

7.3.2 Redundancy

Next, we consider redundancy of the adaptation operators. Recall that an adaptation operator is α is redundant with respect to the DEOL translation

if the first sub-part of the translation is enough to enforce agents to believe what is achieved through the rest of the upgrade: $(\tau(s))^{!C_b(o)} \models \tau(\alpha(s))$ (Definition 3.10 on page 67). For DEOL, it was proved that no adaptation operator is redundant, but that `delete` and `addjoin` are *partially redundant* with respect to agent a: for agent a, the announcement is enough to deduce the beliefs enforced by these adaptation operators.

For ParDEOL, this sub-part of the adaptation operators also includes the forgetting class truth modality $\ominus_{\mathcal{A}\setminus\{b\}} C_b$. This is because on ARG, the class of o is communicated only to check whether the round is a success or not, after which it is discarded. Afterwards, in case it is a failure, an adaptation operator is applied.

Definition 7.13 (Redundancy for ParDEOL)**.** An adaptation operator α is *redundant* if and only if $\forall s$: $(\tau(s))^{!C_b(o);\ominus_{\mathcal{A}\setminus\{b\}} C_b} \models \tau(\alpha(s))$.

We first prove that no adaptation operator is redundant. The reason is that, the announcement $!C_b(o)$ followed by the forgetting class truth modality $\ominus_{\mathcal{A}\setminus\{b\}} C_b$, is not enough for agent b to discard the failing correspondence $\langle C_a, C_b, \sqsupseteq\rangle$ caused by object o. Whereas this correspondence is deleted by any adaptation operator, hence no longer believed in $\tau(\alpha(s))$.

Proposition 7.2 (No redundancy)**.** No adaptation operator is redundant with respect to the translation δ^+.

Proof. Let s be an ARG state and $\langle C_a, C_b, \sqsupseteq\rangle \in A_{ab}$ be the failing correspondence with object o, and consider any adaptation operator α. Then, through the upgrade $!C_b(o);\ominus_{\mathcal{A}\setminus\{b\}} C_b$, agent b acquires neither knowledge nor belief about C_b, while she remains aware of it. However, after applying any adaptation operator α, the correspondence $\langle C_a, C_b, \sqsupseteq\rangle$ is deleted from the alignment. Hence $(\tau(\alpha))^{!C_b(o);\ominus_{\mathcal{A}\setminus\{b\}} C_b} \models B_b(C_a \sqsupseteq C_b)$, but $\tau(\alpha(s)) \not\models B_b(C_a \sqsupseteq C_b)$. Therefore no adaptation operator is redundant. □

But what about partial redundancy? Again, we consider partial redundancy with respect to the upgrade $!C_b(o);\ominus_{\mathcal{A}\setminus\{b\}} C_b$.

Definition 7.14 (Partial Redundancy for ParDEOL)**.** An adaptation operator α is *partially redundant* for agent a if and only if $\tau(\alpha(s)) \models B_a\phi$ implies $(\tau(s))^{!C_b(o);\ominus_{\mathcal{A}\setminus\{b\}} C_b} \models B_a\phi$ for any ARG state s and any ϕ in $\mathcal{L}_{ParDEOL}$.

In DEOL, the announcement $!C_b(o)$ leads agent a to discard the failing correspondence $\langle C_a, C_b, \sqsupseteq\rangle$ and to deduce correspondences involving C_b and classes in her own ontology, exactly what the adaptation operators `delete` and `addjoin` do. However, for ParDEOL, the forgetting class truth modality $\ominus_{\mathcal{A}\setminus\{b\}} C_b$ leads agent a to not be able to deduce these. Therefore, no adaptation operator is redundant nor partially redundant anymore.

Proposition 7.3 (No partial redundancy)**.** No adaptation operator is partially redundant with respect to translation.

Proof. For agent b, this follows directly from redundancy (Proposition 7.2). For agent a, the proof that partial redundancy does not hold follows from the fact that after the announcement $!C_b(o)$, from which agent a comes to know that $C_b(o)$, she discards this again through $\ominus_{\mathcal{A}\setminus\{b\}}C_b$. But then she also discards any correspondence c or $\neg c$ in which C_b occurs, in particular $\neg C_a \sqsupseteq C_b$, which she learned from $!C_b(o)$ but discarded again through $\ominus_{\mathcal{A}\setminus\{b\}}C_b$. However, through any adaptation operator, this correspondence is deleted, leading to agent a believing that $\neg C_a \sqsupseteq C_b$. Therefore for any α: $(\tau(s))^{!C_b(o);\ominus_{\mathcal{A}\setminus\{b\}}C_b} \not\models B_a(\neg C_a \sqsupseteq C_b)$ but $\tau(\alpha(s)) \models B_a(\neg C_a \sqsupseteq C_b)$. Hence no adaptation operator is partially redundant. □

Proposition 7.3 confirms that on ARG, the announcement $!C_b(o)$ is solely used to test the correspondence and decide whether an adaptation operator needs to be applied and it is not used to evolve the alignment.

7.3.3 Completeness

The situation of completeness of the adaptation operators is more interesting to discuss. An adaptation operator α is complete if for all ARG states s: $\tau(\alpha(s)) \models (\tau(s))^{\delta(\alpha)}$ (Definition 3.12 on page 70). In Chapter 3, it was proved that all adaptation operators are incomplete with respect to the DEOL translation δ. This was achieved through the observation that after the announcement $!C_b(o)$, agent a acquires knowledge of this fact and hence $(\tau(s))^{\delta(\alpha)} \models K_a(C_b(o))$, but this knowledge can never be obtained through the adaptation operator as this only affects the alignment, therefore $\tau(\alpha(s)) \not\models K_a(C_b(o))$.

With the new ParDEOL translation δ^+, the situation is different because truth about C_b is forgotten by agent a through $\ominus_{\mathcal{A}\setminus\{b\}}C_b$. This means that this knowledge $K_a(C_b(o))$ does not hold after applying $\delta^+(\alpha)$ to the translation $\tau(s)$ of an ARG state s. Hence, unlike δ, $(\tau(s))^{\delta(\alpha)} \not\models K_a(C_b(o))$.

Let us explain why, still, forgetting class truth is not enough to prove completeness. This is due to the fact that all the upgrades, except $\ominus_{\mathcal{A}\setminus\{b\}}C_b$ that excludes b because $C_b(o)$ is part of her ontolgoy, are *public*. Therefore, even though the agents (except b) discard $C_b(o)$ and therefore what was learned from $!C_b(o)$ is unlearned, still *each* agent acquires the belief of the new correspondence that is added through the adaptation operators $\alpha \in \{\texttt{add}, \texttt{addjoin}, \texttt{refine}, \texttt{refadd}\}$, if they are aware of the classes in this correspondence. This awareness cannot be ruled out as there might always be other agents such that these classes also appear in correspondences in

her alignment. However, we can prove that all the adaptation operators are complete for ARG states consisting of only two agents.

Before we prove this, the effect of one difference between logical and adaptive agents as discussed in Chapter 4 needs to be discussed with respect to completeness. The difference arising out of the ability of logical agents to combine beliefs and the inability of adaptive agents to combine alignments. This difference is certainly still present in the translation δ^+: $\big(B_a(C(x)) \wedge B_a(C \sqsubseteq D)\big) \rightarrow B_a(D(x))$ and $\big(B_a(C \sqsubseteq C') \wedge B_a(C' \sqsubseteq D)\big) \rightarrow B_a(C \sqsubseteq D)$ also hold for ParDEOL as long as agents are aware of C, D, C' – which is the case for all classes occurring in correspondences of the alignment. This does not affect completeness because these beliefs are combined in both $(\tau(s))^{\delta(\alpha)}$ and $\tau(\alpha(s))$, and the beliefs that agents can combine to reach new beliefs are the same in both situations.

Proposition 7.4 (Completeness)**.** All adaptation operators are complete *for ARG states consisting of two agents* with respect to the translation δ^+.

Proof. Let s be an ARG state for a set of agents $\mathcal{A} = \{a, b\}$ and let α an adaptation operator for ARG. We need to show that $\tau(\alpha(s)) \models (\tau(s))^{\delta^+(\alpha)}$ (Definition 3.12 on page 70).

Assume that $(\tau(s))^{\delta^+(\alpha)} \models \phi$ for some $\phi \in \mathcal{L}_{ParDEOL}$. Then if ϕ is a non-epistemic formula, also $(\tau(s)) \models \phi$ because the only hard information gained through $\delta^+(\alpha)$ is $C_b(o)$ (from the announcement $!C_b(o)$), which is also discarded through $\ominus_{\{a\}} C_b$ in $\delta^+(\alpha)$. The forgetting modality $\ominus_{\{a\}} C_b$, even if it only alters the knowledge and beliefs of agent a, ensures that $C_b(o)$ is not true globally: in the aware cell of agent a, there are worlds such that $C_b(o)$ is false, making $C_b(o)$ not to hold globally. Hence $(\tau(s))^{\delta^+(\alpha)} \not\models C_b(o)$ and therefore $(\tau(s)) \models \phi$. Then also $(\tau(\alpha(s))) \models \phi$ because α only affects the alignment between agents but not truth.

If ϕ is an epistemic formula, there are two cases: (1) $\phi = K_i\psi$ or (2) $\phi = B_i\psi$ for some $i \in \mathcal{A}$.

(1) Assume that $(\tau(s))^{\delta^+(\alpha)} \models K_i\psi$. Then, for the same reasoning as before, it must be that $(\tau(s)) \models K_i\psi$: the knowledge agent a acquires through the announcement $!C_b(o)$ of $\delta^+(\alpha)$, $K_a(C_b(o))$, is immediately discarded by $\ominus_{\{a\}} C_b$ and no other knowledge is gained through $\delta^+(\alpha)$. Then, because α does not affect the ontologies of agents, it also holds that $\tau(\alpha(s)) \models K_i\psi$.

(2) Assume that $(\tau(s))^{\delta^+(\alpha)} \models B_i\psi$. Then, either (i) ψ is a correspondence CRD occurring in the conservative upgrade of $\delta^+(\alpha)$, or (ii) it is obtained from combining this correspondence with knowledge or belief that already held in $\tau(s)$. In both cases, this correspondence is also believed in $\tau(\alpha(s))$ by the agent: for (i) this is true because if $\uparrow(CRD)$ occurs in $\delta^+(\alpha)$ then CRD is the correspondence added by α to the alignment between agents a and b

at s, hence it will be translated to belief by τ. For (ii) this holds because (i) ensures that the only belief added in the logical model is also added by the adaptation operator, while the knowledge and other beliefs of agents stay the same, and agents can combine their knowledge and beliefs in the same ways in $\tau(\alpha(s))$ and $\tau(s)^{\delta^+(\alpha)}$. Hence, $\tau(\alpha(s)) \vDash B_i\psi$

Therefore, for any $\phi \in \mathcal{L}_{ParDEOL}$ such that $(\tau(s))^{\delta^+(\alpha)} \vDash \phi$ it holds that $(\tau(\alpha(s))) \vDash \phi$. Hence $\tau(\alpha(s)) \vDash (\tau(s))^{\delta^+(\alpha)}$. □

The completeness result cannot be extended to ARG states with more than two agents because the conservative upgrade causes all agents to believe the new correspondence, while this can never be part of the alignment.

Proposition 7.5 (Incompleteness)**.** For any ARG state s with $|\mathcal{A}| > 2$, `delete` is complete with respect to the translation δ^+. The other adaptation operators are incomplete with respect to δ^+.

Proof. The adaptation operator `delete` is complete because it does not add any correspondence and all agents that learn $C_b(o)$ through $!C_b(o)$ discard it again through $\ominus_{\mathcal{A}\setminus\{b\}}$.

The other adaptation operators are incomplete because the added correspondence by the operator will be believed by *all* agents that are aware of the classes in this correspondence, also $c \notin \{a, b\}$. However, these correspondences themselves are not part of their alignments in ARG, because these alignments only include correspondences between classes in c's ontology and other agents' ontologies.

□

Completeness of the adaptation operators for ARG states with two agents, with respect to the translation δ^+, shows that awareness and the forgetting class truth modality indeed bring the logical model of ARG closer to the original game. However, it also shows the need for private communication in order to be extended to ARG states with more than two agents because public communication causes all the agents to update their knowledge and beliefs with the upgrades. Replacing the announcement $!C_b(o)$ by a private announcement $!_{\{a\}}C_b$, the forgetting class truth $\ominus_{\mathcal{A}\setminus\{b\}}$ and finally the conservative upgrade $\uparrow c$ for a correspondence c by a private conservative upgrade to only agents a and b could therefore benefit the logical model and may lead to completeness for all ARG states.

Finally, the proofs of correctness, redundancy and completeness for $|\mathcal{A}| = 2$ emphasize that the lack of formal properties can be interpreted in two ways: either the adaptive agents are sub-logical or the logical model in DEOL is insufficient to model their behavior. ParDEOL contributes to the second interpretation.

7.4 The Evolution of Awareness in ARG

With a formal notion of awareness, we can explore how awareness evolves on ARG. Recall that the awareness of an agent a at a world w of a model $\mathcal{M}$ can be defined with respect to the aware cell of an agents, i.e. $AW_a(w) = \bigcup_{w' \in ||w||_a} Dom(V_{w'})$. For a pointed model $\langle \mathcal{M}, w \rangle$, $AW_a(\mathcal{M})$ can be used to denote $AW_a(w)$ when w is clear from the context, and for a theory T, $AW_a(T)$ can be used to denote the intersection of the awareness of a in all pointed models $\langle \mathcal{M}, w \rangle$ of T, i.e. $AW_a(T) = \bigcap_{\langle \mathcal{M}, w \rangle \models T} AW_a(w) = \bigcap_{w' \in ||w||_a \text{ s.t. } \langle \mathcal{M}, w \rangle \models T} Dom(V_{w'})$. When it is not specified otherwise these pointed models will be assumed to be ParDEOL models.

First, awareness in ParDEOL is confined within full awareness in DEOL. That is, the awareness of an agent at a pointed ParDEOL model of the translation of an ARG state is included in the awareness of that agent at the DEOL model.

Proposition 7.6 (Awareness is confined)**.** Agent awareness in ParDEOL is confined within full awareness (in DEOL): $\forall s, \forall a \in \mathcal{A}$: if $\langle \mathcal{M}^+, w^+ \rangle$ is a ParDEOL model of $\tau(s)$ and $\langle \mathcal{M}, w \rangle$ is a DEOL model of $\tau(s)$ then $AW_a(\mathcal{M}^+) \subseteq AW_a(\mathcal{M})$.

Proof. Let s be an ARG state and $\tau(s)$ be its logical translation. Furthermore, let $\langle \mathcal{M}^+, w^+ \rangle$ and $\langle \mathcal{M}, w \rangle$ be ParDEOL and DEOL models of $\tau(s)$, respectively. Then in $\langle \mathcal{M}^+, w^+ \rangle$, agents know their own ontology and believe the alignments between their ontology and other agents' ontologies. Additionally, this encompasses the awareness of agents, i.e. $AW_a(\mathcal{M}^+) = sig(\mathcal{O}_a) \cup \{C_b \mid \langle C_a, C_b, R \rangle \in \bigcup_{b \in \mathcal{A} \setminus \{a\}} A_{ab}\}$. But in $\langle \mathcal{M}, w \rangle$, agents have full awareness of all the classes in their own *and* other agents' ontologies. In fact, they are aware of the set of propositions, i.e. $AW_a(\mathcal{M}) = P$. Hence $AW_a(\mathcal{M}^+) \subseteq AW_a(\mathcal{M})$. □

Now, let us consider how awareness evolves in ParDEOL. This can be explored in two different ways: (1) how does awareness evolve through applying the adaptation operator α and (2) how does it evolve through the logical dynamics $\delta^+(\alpha)$ that are translations of α in ParDEOL. More specifically, the first corresponds to comparing awareness of agents on $\tau(s)$ with $\tau(\alpha(s))$, and the second corresponds to comparing this with $(\tau(s))^{\delta^+(\alpha)}$. See Figure 7.5.

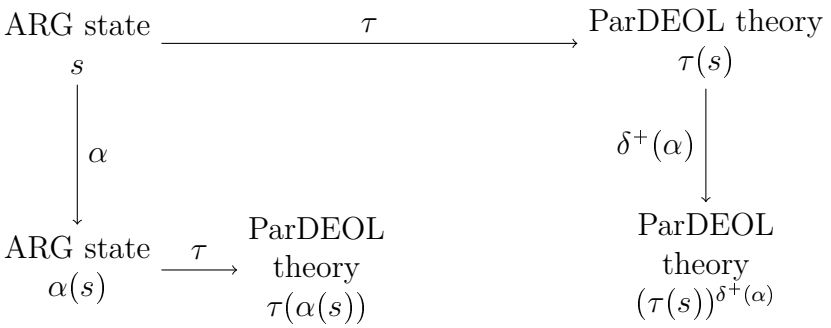

Figure 7.5: Relations between ARG scenarios and ParDEOL theories via the translation τ, and between adaptation operators α and ParDEOL dynamics $\delta^+(\alpha)$.

First, let us take a look at δ^+. In ParDEOL, agent awareness may only increase through application of logical dynamics corresponding to all adaptation operators.

Proposition 7.7 (Awareness increases via the logical dynamics $\delta^+(\alpha)$)**.** Let s be an ARG state and α be any permitted adaptation operator applied to a failing correspondence $\langle C_a, C_b, R\rangle$ with object o, then $\forall a \in \mathcal{A}$: $AW_a((\tau(s))^{\delta^+(\alpha)}) \supseteq AW_a(\tau(s))$.

Proof. By definition of $\delta^+(\alpha)$, agents either preserve their awareness (for example for `delete`), or raise their awareness. But awareness cannot decrease since no forgetting class awareness modality is present in the translation of adaptation operators. Therefore $AW_a((\tau(s))^{\delta^+(\alpha)}) \supseteq AW_a(\tau(s))$ for all agents a. □

The different adaptation operators raise different agent's awareness. For `delete`, no awareness is raised and both agents preserve their awareness at $\tau(s)$, `add`, `addjoin` and `refine` raise the awareness of one of the agents (`add` and `addjoin` that of agent b, and `refine` that of agent a – because the other agent is already aware of this class as it appears in her ontology) and `refadd`, as the combination of `addjoin` and `refine`, raises awareness of both agents.

The same is not true for ARG: through application of an adaptation operator, awareness may actually decrease. This is in particular the case for the adaptation operator `delete`. This operator deletes the failing correspondence from the alignment and therefore, if the classes of this correspondence do not occur in another correspondence, the ParDEOL model of the logical translation of the resulting state excludes these classes from the awareness of agents.

Proposition 7.8 (Awareness may decrease via the adaptation operator α)**.** There is an ARG state s with permitted adaptation operator α and an agent a such that $AW_a(\tau(\alpha(s))) \not\subseteq AW_a(\tau(s))$.

Proof. Let s be an ARG state and A_{ab} be the alignment in this state of agents a and b. Now assume that the failing correspondence is $C_a \sqsupseteq C_b$ with object o and that these classes do not appear in other correspondences of the alignment. Finally, let the applied adaptation operator be `delete`. This means that $C_a \sqsupseteq C_b$ is deleted from A_{ab} in $\texttt{delete}(s)$. Moreover, this means that the classes C_a and C_b do no longer continue to appear in the alignment of the new state $\texttt{delete}(s)$. Therefore, in the translation $\tau(\texttt{delete}(s))$, agent a does not hold any belief, nor knowledge, about C_b and agent b not about C_a. Hence in the ParDEOL model of this state, $C_b \notin AW_a(\tau(\texttt{delete}(s)))$ and $C_a \notin AW_b(\tau(\texttt{delete}(s)))$. However, since in s, $C_a \sqsupseteq C_b$ did belong to the alignment A_{ab}, $C_b \in AW_a(\tau(s))$ and $C_a \in AW_b(\tau(s))$. Thus $AW_i(\tau(\alpha(s))) \not\subseteq AW_i(\tau(s))$ for $i \in \{a, b\}$. □

Proposition 7.8 illustrates that adaptive agents may not only forget truth, but may also forget awareness, and Proposition 7.7 shows that this does not hold for logical agent. The question therefore may be why not use $-_{\mathcal{A}\setminus\{b\}}C_b$ in the translation δ^+ of the adaptation operators (Definition 7.12 on page 145)? The answer is, that this may violate the translation $\tau(s)$ and its ParDEOL models because the situation used in the proof of Proposition 7.8 is very specific: if agents have another correspondence involving the class C_b, forgetting awareness of this class causes agents to stop believing this correspondence, hence the alignment on ARG is no longer translated to beliefs in ParDEOL.

From Propositions 7.7 and 7.8 we can conclude that ParDEOL models correspond to ARG but retain acquired awareness.

Proposition 7.9 (ParDEOL preserves awareness)**.** Let s be an ARG state and α any permitted adaptation operator applied to a failing correspondence $\langle C_a, C_b, R\rangle$ with object o, then $\forall a \in \mathcal{A}$: $AW_a(\tau(\alpha(s))) \subseteq AW_a((\tau(s))^{\delta^+(\alpha)})$.

Proof. Consider an ARG state s and let α be any permitted adaptation operator applied to failing correspondence $\langle C_a, C_b, R\rangle$ with object o. Then the classes agents may become aware of by application of α ($msc_a(C_a)$ for `add`, $mscc_a(o, C_a)$ for `addjoin`, $\{C'_b\}$ for `refine` and the latter two for `refadd`) are exactly those classes of which awareness is raised by the dynamics of $\delta^+(\alpha)$. And because awareness may not decrease through δ^+, for all agents a: $AW_a(\tau(\alpha(s)) \subseteq AW_a((\tau(s))^{\delta^+(\alpha)})$. □

Proposition 7.10 (Incomplete Awareness)**.** There is an ARG state s and adaptation operator α that is applied to a failing correspondence $\langle C_a, C_b, R\rangle$ with object o such that $AW_a(\tau(\alpha(s))) \not\supseteq AW_a((\tau(s))^{\delta^+(\alpha)})$.

Proof. Because awareness may decrease via α (Proposition 7.8), but ParDEOL preserves awareness (Proposition 7.9) we have there are situations such that $AW_a(\tau(\alpha(s))) \not\ni AW_a((\tau(s))^{\delta^+(\alpha)})$. □

In Figure 7.6, the propositions are visualized with respect to the diagram of translations τ and δ^+.

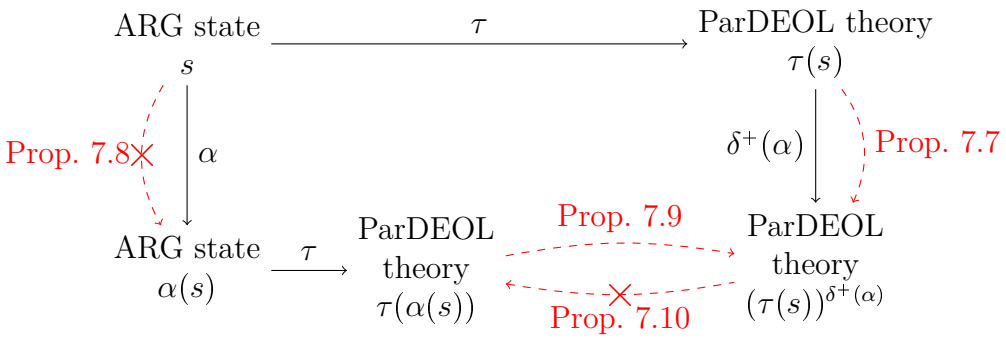

Figure 7.6: Relations between ARG scenarios and ParDEOL theories via the translation τ, and between adaptation operators α and ParDEOL dynamics $\delta^+(\alpha)$. The red dashed arrows indicate the inclusion of awareness of the source in that of the target (and when it is crossed, it means that it does not hold).

Despite the fact that awareness may decrease on ARG through application of the adaptation operators, experiments have shown that agents converge to a state with successful communication and stable alignments [46, 49]. This suggests that agents do not need full awareness to communicate successfully, they only need to be aware of a subset of the vocabularies used by other agents that is sufficient for successful communication. Consider the following example.

Example 7.1. Let s^* be the ARG state drawn in Figure 7.7 where the alignment A_{ab} is the set of correspondences $\langle C_a^l, C_b^l, R\rangle$ such that $C_a^l \in \mathcal{O}_a$ and $C_b^l \in \mathcal{O}_b$ are "leaf classes" (they have no subclass) and R is the truthful relation holding between them, corresponding to the reference alignment. Then any object drawn will reach a success in ARG and for all $i \in \mathcal{A}, \forall\alpha$: $AW_i(\tau(s^*)) = AW_a(\tau(\alpha(s^*)))$, but agents do not have full awareness.

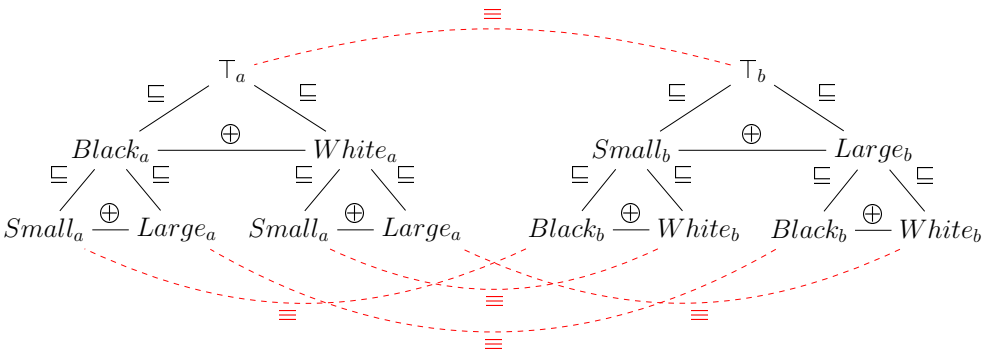

Figure 7.7: An ARG state in which the alignment consists of all truthful correspondences between leaf classes.

When ARG reaches a stable state, a state in which every round of ARG is a success, awareness in the logical model is also stable. This is because no more adaptation operator is applied, and hence also no awareness is raised. Furthermore, also the knowlege and beliefs of agents are stable in this case: even though the upgrades $!C_b(o); \ominus_{\mathcal{A}\setminus\{b\}}$ continue to be applied, this does not cause agents to acquire information: through $!C_b(o)$ agents (except b) learn that $C_b(o)$ but this is immediately forgotten because of $\ominus_{\mathcal{A}\setminus\{b\}} C_b$.

7.5 Conclusion

In this chapter, an alternative translation of ARG is provided in ParDEOL, making use of raising class awareness and forgetting class truth modalities. This enables agents to use different vocabularies to express their knowledge and beliefs and extend their vocabularies via raising awareness modalities when encountering new classes, overcoming full vocabulary awareness.

With this new translation, it was shown that, unlike the translation in DEOL, the adaptation operators are correct, not (partially) redundant and complete for ARG states of two agents with respect to the translation in ParDEOL. This is in particular the case because the forgetting class truth modality causes the agents to discard the classification of the object that was used to test the alignment. As a result, the DEOL translation can be considered insufficient to model ARG, and the ParDEOL translation as an improvement, bringing the logical model closer to the original game. However, the proof that `add` is correct for ParDEOL also shows that adaptive agents benefit from considering logical adaptation operators. Furthermore, incompleteness for ARG states with more than two agents, compared to

completeness for two agents, suggests the need for switching to private dynamic upgrades, in particular for the conservative upgrade introducing the new correspondence.

Awareness, raising awareness and forgetting therefore bring the logical model of ARG closer to the original game in two ways: (1) by enabling distinct and dynamic vocabularies, and (2) by letting agents discard evidence. This emphasizes that indeed the lack of formal properties can be interpreted and addressed in two ways: by giving the adaptive agents more logical power, or by bringing the logical model closer to their behavior.

Chapter 8

Conclusion

> Je ne sais pas où je vais, oh ça
> je l'ai jamais bien su
> Mais si jamais je le savais, je
> crois bien que je n'irai plus
>
> Le Rue Ketanou, *Où je vais*

In this thesis, a logical model for the Alignment Repair Game (ARG) is introduced in order to examine its formal properties. This proved more challenging than thought and led to identify three differences between adaptive agents and logical agents. We then introduced awareness to bring the logical model closer to the original game. In this chapter, we give a summary of our contribution and we present the perspectives for ARG as well as beyond ARG.

Summary

This thesis started with the question: what are the formal properties of the adaptation operators, are they logically correct, complete or redundant? To answer these questions, a logical model was introduced for ARG. This model was based on Dynamic Epistemic Ontology Logic (DEOL), which extends Dynamic Epistemic Logic (DEL) with object classifications and class relations from a simple Description Logic language. A translation from ARG to DEOL was defined that maps agents' ontologies and alignments to knowledge and beliefs, respectively, and translates adaptation operators to announcements and conservative upgrades. The translation was proven faithful (it preserves consequences of ontologies and alignments, and the knowledge is the same if ARG is consistent) and ARG state preserving.

The translation enabled to define the formal properties (correctness, completeness and redundancy) of the adaptation operators. We then proved that all adaptation operators but `add` are formally correct, all adaptation operators are incomplete and `delete` and `addjoin` are partially redundant. With these results, this thesis bridges a very practical implementation of adaptive agents used in simulations with a dynamic epistemic model of logical agents. In spite of, or because of, the simplicity of ARG, this revealed more challenging than expected.

This led us to identify three differences between adaptive agents and their logical model (logical agents): (1) adaptive agents reason locally while logical agents reason globally, (2) logical agents share a fixed vocabulary, preventing them from using heterogeneous knowledge representations like adaptive agents, and (3) adaptive agents are unable to remember individual cases because they focus on general knowledge, whereas logical agents cannot discard these.

To reduce these differences, the assumption of full vocabulary awareness was dropped, which holds on DEL and DEOL. This assumption prevents agents from using distinct and dynamic vocabularies to represent their knowledge and beliefs. To respect heterogeneity between agents, a notion of awareness was introduced based on partial valuation functions and weakly reflexive relations: Partial Dynamic Epistemic Logic (ParDEL) was born. ParDEL enables to distinguish *uncertain* agents, agents that are aware of a proposition but do not know the truth value, from *unaware* agents, agents that do not consider the proposition at all. The properties of awareness were motivated and formalized, and the syntax and semantics of ParDEL were defined. Modalities for raising awareness were introduced for ParDEL that allow agents to extend their vocabularies when encountering new terms, either from the environment or through interaction with other agents. We proved that these modalities are disconnected from truth: raising awareness does not disclose truth values.

With a formal notion of awareness, the next step was to define forgetting modalities. These modalities were motivated by the difference between adaptive agents and logical agents. More precisely, that adaptive agents discard evidence in favour of general knowledge, whereas logical agents cannot forget. Two forgetting modalities were introduced: forgetting awareness and forgetting truth. The relation between the two types was explored and it was shown that the models obtained from applying forgetting awareness and from applying forgetting truth followed by raising awareness are bisimilar. Additionally, we proved that forgetting awareness implies forgetting truth, making it not the exact converse of raising awareness.

Awareness, raising awareness and forgetting were introduced as an at-

tempt to bring the logical model of ARG closer to the original version of the game. To confirm this, we defined a new translation, from ARG to ParDEOL, and re-examined the formal properties of the adaptation operators with respect to this translation. We proved that the adaptation operators are correct, complete when the ARG state consists of two agents and no longer (partially) redundant, confirming that indeed DEOL is insufficient to capture the behavior of adaptive agents and an alternative is needed. This shows that DEOL, and therefore DEL, is not yet a good logical model for cultural knowledge evolution. Finally, with the notion of awareness, we explored how awareness evolves through playing ARG and we proved that agents become aware only of the sub-vocabularies necessary to succeed in the game, which may not be the full vocabularies.

Contribution and Perspectives

The logical model of ARG presented in this thesis helps to understand the properties of adaptation operators: correctness, completeness and redundancy. However, it also proved challenging to model an experimental cultural knowledge evolution game in logic, leading us to identify three differences between adaptive agents and logical agents. We addressed two of them by introducing awareness and modalities for raising awareness and forgetting, which brought the logical model closer to the original ARG game and enabled us to prove the formal properties.

Beyond the specific case of ARG, the contribution of the logical model presented in this thesis is two-fold: (1) it shows how cultural knowledge evolution can be assessed theoretically in logic and which challenges are faced in the logic by doing so, and (2) it introduces an independent model of awareness enabling agents to use different and dynamic vocabularies to express their knowledge.

The logical model of ARG more broadly defines a theoretical model of cultural knowledge evolution. In particular, it provides a specific methodology to translate ontologies and alignments to agents' knowledge and beliefs, respectively, and to capture communication and interaction between agents with dynamic upgrades – while respecting heterogeneity and autonomy. This revealed the need for adjusting the logic to enable agents to use their own, private vocabularies that are shared whenever necessary for achieving successful communication. As such, it is in line with autonomous evolution of heterogeneous knowledge in which agents do not wait for knowledge to be perfect before using it and agents, or societies of agents, cannot be inter-

rupted to upgrade their knowledge. Furthermore, we showed how to define formal properties that could be satisfied. Of course, these are properties of both the game and the translation, hence one must be cautious about what the translation preserves (faithfulness). As such, this methodology can be applied to other multi-agent knowledge evolution experiments to assess them, bridging a very practical implementation of adaptive agents used in simulations with a dynamic epistemic model of logical agents.

Moreover, this thesis introduces a model of awareness, enabling to model situations in which agents use different and dynamic vocabularies to express their knowledge representations and raising awareness is disconnected from learning truth. This is a stand alone contribution, independent from the ARG modeling and can be used to capture private dynamic upgrades in logic. Such private upgrades typically violate the properties of DEL models, in particular the reflexivity condition, and are therefore usually excluded. Because ParDEL only requires weak reflexivity, this argument does not apply: reflexivity cannot be violated by private upgrades. Therefore, ParDEL may be a suitable framework to model private communication and interaction between a group of agents.

ARG covers only relatively simple ontologies and alignments, for example they do not involve roles, and extending the logical model to capture these would quickly move the model into First Order Logic. However, even with the restrictions on ontologies and alignments of ARG, quite some challenges arose that we have tried to overcome by bringing the logical model closer to ARG, through introducing awareness and forgetting operations. Yet, there still remain differences between adaptive agents and logical agents. In particular, the difference concerning local versus global reasoning: adaptive agents use their alignments one by one, while logical agents combine all their beliefs, coming from the translation of the alignments. A perspective is to address this difference. Incorporating the *society-of-minds* approach [26, 43, 51, 78] could benefit the logical model of ARG and bring it yet closer to its original version. In this approach, agents are viewed as sets of different belief clusters. Each of these clusters is locally consistent, but global inconsistencies are not ruled out. Then, depending on the situation, agents 'choose' a cluster with which to reason, temporarily forgetting about the others. This can be compared to ARG, in which the agents use alignments one by one, depending on the agent it is interacting with. It may therefore enable logical agents to use their beliefs, which are translated from the alignments, one by one and thus preventing them to combine all their beliefs together. This would be more faithful, in particular concerning *strict belief adherence* (Section 3.2.2 on page 58).

There are also perspectives of the partial logics introduced. Firstly, they can be linked to other approaches to model awareness that do not use partial valuations/interpretations, in particular to work considering higher order knowledge and beliefs with respect to awareness, for example in which agents can reason about unawareness of other agents [2, 57]. Since unawareness cannot be captured directly in the language when using partial valuations/interpretations, it is an open question whether a statement such as "agent a knows that agent b is not aware of p" can be formalized and how it compares.

Secondly, in DEL, reduction axioms are used to 'reduce' formulas with dynamics, such as announcements, to a formula without dynamics recursively. This is used to reduce expressivity, soundness and completeness to the case of modal logic. In the logics of awareness discussed here, these reduction axioms do not translate directly. This is because preconditions are used in a different way to determine the product update of event models and epistemic models: the preconditions are used to select worlds that "do not falsify it", i.e. they either verify the precondition, or the precondition is undefined. Since undefinedness cannot be captured on the partial logics, this complicates reducing formulas with raising awareness modalities to formulas without. Perhaps a three-valued approach would be useful for this purpose and could be used to define an axiomatization for ParEAL and eventually ParDEOL.

Thirdly, we have defined (private) raising and forgetting propositional/class awareness modalities and used these to introduce a (private) forgetting propositional/class truth modality. However, when considering classes on ParDEOL, there is another option for agents to forget:*forgetting classifications* $\ominus C(o)$. This can be interpreted as forgetting that object o belonged to C but not forgetting C in its entirety. This could be achieved by duplicating all the worlds in which a class C is defined, and adjusting the interpretation in this duplicated world in such a way that if o belonged to the interpretation of C in the initial world, it will not in the duplicated world, and vice versa. Such a modality would then preserve all the other classifications of objects to C, but only forgets whether or not the object o belonged to C. This could eventually bring the logical model closer to ARG as such a modality exactly reverses the announcement $!C_b(o)$.

Another perspective for ARG is to characterize expansion and relaxation [49]. These features allow agents to introduce new random correspondences (expansion) or to use shadowed correspondences (relaxation) to improve their alignments, leading to better alignments as the experiments have shown [49]. With a formal notion of awareness, expansion could be cap-

tured by raising awareness and conservative upgrades. Whether awareness is enough to capture relaxation is an open question.

The Alignment Repair Game, and more generally cultural knowledge evolution, can also be linked to belief revision. Indeed cultural knowledge evolution needs belief representation and belief revision when beliefs do not reveal adapted. Although we have only studied ARG under the light of belief contraction, belief revision has been thoroughly considered in DEL [8,12]. Therefore, this may be used to attack the problem of modeling cultural knowledge evolution in logic more generally.

Appendix A

Proofs of Faithfulness

Proof of Proposition 3.1.
(Equation 3.8)

We consider two cases: (1) $\tau(s)$ has a model and (2) $\tau(s)$ has no model. Whenever (2), Equation 3.8 trivially holds: every model of $\tau(s)$ (which are zero) makes $K_a\phi$ true and therefore $\mathcal{O}_a \models \phi \Rightarrow \tau(s) \models K_a\phi$.

Otherwise, when $\tau(s)$ has a model, for any $\psi \in \mathcal{O}_a$, by Definition 3.5 (page 51) of the translation, $K_a\psi$ is an axiom of $\tau(s)$. This means that for any model $\langle \mathcal{M}, w\rangle$ of $\tau(s)$ and any world v accessible for a from w, $\mathcal{M}, v \models \psi$. Now assume that $\tau(s) \not\models K_a\phi$ for some ϕ. This means that there exists a model $\langle \mathcal{M}, w\rangle$ of $\tau(s)$ and a world v reachable for a from w in which $\mathcal{M}, v \not\models \phi$. In such a case, the interpretation I_v at world v of $\mathcal{M}$ would be a model of O_a (because, by the translation, it satisfies all axioms of O_a), thus $O_a \not\models \phi$. Hence, the contraposition holds: if $\mathcal{O}_a \models \phi$ then $\tau(s) \models K_a\phi$.
(Equation 3.9)

Again, we consider the two cases: (1) $\tau(s)$ has a model and (2) $\tau(s)$ has no model, and whenever (2), Equation 3.9 trivially holds: every model of $\tau(s)$ (which are zero) makes $B_a\gamma$ true and therefore $A_{ab} \models \gamma \Rightarrow \tau(s) \models B_a\gamma$.

Otherwise, when $\tau(s)$ has a model, for any $\psi \in \mathcal{O}_a$, $K_a\psi \in \tau(s)$ (Definition 3.5 on page 51), so $\tau(s) \models B_a\psi$. In addition, for any $\gamma' \in A_{ab}$, $B_a\gamma' \wedge B_b\gamma'$ is an axiom of $\tau(s)$ (Definition 3.5 on page 51). This means that for any model $\langle \mathcal{M}, w\rangle$ of $\tau(s)$ and any world v most plausible for a from w, $\mathcal{M}, v \models \psi$ and $\mathcal{M}, v \models \gamma'$. Now assume that $\tau(s) \not\models B_a\gamma$ for some correspondence γ. This means that there exists a model $\langle \mathcal{M}, w\rangle$ of $\tau(s)$ and a world v that is considered most plausible for a from w in which $\mathcal{M}, v \not\models \gamma$. In such a case, the interpretation I_v of $\mathcal{M}$ at v would be an extension of a model of O_a to $\mathcal{C}_a \cup \mathcal{C}_b$ (because by the translation it satisfies all axioms of O_a and all correspondences of A_{ab}), thus $A_{ab} \not\models_a \gamma$. Hence, the contraposition holds: if $A_{ab} \models_a \gamma$ then $\tau(s) \models B_a\gamma$. □

Proof of Lemma 3.1.
We need to show that $\forall a \in \mathcal{A}$, $\forall b \in \mathcal{A}\backslash\{a\}$, $\forall \phi_a \in \mathcal{O}_a$, $\forall \gamma \in A_{ab}$: $\mathcal{M}^s, w^s \models K_a(\phi_a) \wedge B_a(\gamma)$. Hence, that $\forall \phi_a \in \mathcal{O}_a$: $\mathcal{M}^s, w^s \models \phi_a$ and $\mathcal{M}^s, w_a \models \phi_a$ and (2) $\forall \gamma \in A_{ab}$: $\mathcal{M}^s, w_a \models \gamma$.

$[\mathcal{M}^s, w^s \models \phi_a]$ At w^s, the interpretation is such that it assigns to each class C the standard interpretation $\hat{I}_a$ for agent a such that $C \in \mathcal{C}_a$. And because by construction the standard interpretation $\hat{I}_a$ for $\mathcal{O}_a$ satisfies each ϕ_a in $\mathcal{O}_a$: $\mathcal{M}^s, w^s \models \phi_a$.

$[\mathcal{M}^s, w_a \models \phi_a \wedge \gamma]$ At w_a, the interpretation is I'_a, an extension of I_a of $\mathcal{O}_a$ to $\bigcup_{a\in\mathcal{A}} \mathcal{C}_a$ such that $I_a \models_a A_{ab}$ for each $b \in \mathcal{A}\backslash\{a\}$. This means that (1) I'_a satisfies all ϕ_a in $\mathcal{O}_a$, because I_a does, and (2) I'_a satisfies all the alignments involving a. Hence, $\forall \phi_a \in \mathcal{O}_a$: $\mathcal{M}^s, w_a \models \phi_a$, and $\forall \gamma \in \bigcup_{b\in\mathcal{A}\backslash\{a\}} A_{ab}$: $\mathcal{M}^s, w_a \models \gamma$.

Therefore each agent knows her ontology and believes her alignments and so $\langle \mathcal{M}^s, w^s \rangle$ is a model of $\tau(s)$. □

Proof of Proposition 3.2.
(if) Assume that s is an ARG state for a set of agents $\mathcal{A}$ with locally consistent alignments. This means that, for each agent a there is a local model I_a for $\mathcal{O}_a$ of the alignments involving a, i.e. of $\bigcup_{b\in\mathcal{A}\backslash\{a\}} A_{ab}$. Thus there is an extension I'_a of I_a that locally satisfies all $\gamma \in \bigcup_{b\in\mathcal{A}\backslash\{a\}} A_{ab}$ and all $\phi_a \in \mathcal{O}_a$. We need to show that $\tau(s)$ has a model. Let $\langle \mathcal{M}^s, w^s \rangle$ be as defined in Definition 3.6 (page 55), then, by Lemma 3.1, $\langle \mathcal{M}^s, w^s \rangle$ is a model of $\tau(s)$.
(only if) Assume that s is an ARG state for a set of agents $\mathcal{A}$ such that s does not have locally consistent alignments. Thus, there is an agent $a \in \mathcal{A}$ for which there exists no local model I_a for $\mathcal{O}_a$ of the alignments involving a. In other words, for each model I_a of $\mathcal{O}_a$, there exists no extension I'_a that satisfies A_{ab} for each $b \in \mathcal{A}\backslash\{a\}$. Now suppose, towards a contradiction, that there is a model $\langle \mathcal{M}, w \rangle$ of $\tau(s)$. Then, by the translation, it must be that $\forall \gamma \in \bigcup_{b\in\mathcal{A}\backslash\{a\}} A_{ab}$: $\mathcal{M}, w \models B_a\gamma$. Thus, each correspondence in an alignment involving a must be true at any of the most plausible worlds v for a from w. In addition, $\forall \phi \in \mathcal{O}_a$, again by the translation, it holds that $\mathcal{M}, w \models K_a\phi$. I.e. ϕ should be true in all accessible worlds for a from w, and, in particular, at v. This means that the interpretation I_v at world v must be a model of $\mathcal{O}_a$ and therefore, since s is not locally consistent, no extension of I_v exists that locally satisfies all the alignments involving a. But then it cannot be the case that all correspondences $\forall \gamma \in \bigcup_{b\in\mathcal{A}\backslash\{a\}} A_{ab}$ are true at v. Hence, we arrive at a contradiction. Thus there is no such model. □

To prove Proposition 3.3, first a property of standard interpretations

(Definition 2.5 on page 14) is proven: for two different classes C, D, if the standard interpretation of D is an extension of that of C, the reverse cannot hold. This is true because ARG ontologies are constructed as dichotomic trees.

Lemma A.1. For each locally consistent ARG state s and for each standard model of $\tau(s)$, for every ARG ontology $\mathcal{O}_a$ over signature $\langle \mathcal{C}_a, \mathcal{D} \rangle$ and for every two classes $C, D \in \mathcal{C}_a$: if $C^{\hat{I}} \subseteq D^{\hat{I}}$ and $C \neq D$ then $D^{\hat{I}} \nsubseteq C^{\hat{I}}$.

Proof. By Definition 3.6 (page 55), any class $C, D \in \mathcal{C}$ are different and their interpretation is not empty. Hence, there is at least an object $o \in \mathcal{D}$ such that $o \in C^{\hat{I}}$ and, given that $C^{\hat{I}} \subseteq D^{\hat{I}}$, $o \in D^{\hat{I}}$. By definition of ARG ontologies (Definition 2.2(4) on page 12), for any $o \in \mathcal{D}$, there is a unique $C^o \in \mathcal{C}$ such that $C^o(o) \in \mathcal{O}_a$. Hence, $o \in C^{\hat{I}}$ means that either (a) $C = C^o$ or (b) $\exists C', C'' \in \mathcal{C}$; $C' \sqsubseteq C \in \mathcal{O}$, $C'' \sqsubseteq C \in \mathcal{O}$, and $o \in C''^{\hat{I}}$ and $o \notin C'^{\hat{I}}$ (because $C' \oplus C'' \in \mathcal{O}_a$, Definition 2.2(3) on page 12). By induction, this means that there is a chain of classes $C_1, C_2, \ldots, C_n \in \mathcal{C}_a$ with $C_i \sqsubseteq C_{i+1} \in \mathcal{O}_a$, $C_1 = C$ and $C_n = C^o$. The same holds true for D. Moreover ([19]), there is a unique chain $C_1, C_2, \ldots, C_n \in \mathcal{C}_a$ with $C_i \sqsubseteq C_{i+1} \in \mathcal{O}_a$, $C_1 = C^o$ and $C_n = \top$. Since this also holds true for C and D, this means that they both belong to this chain ($\exists i, j \in [1, n]$ such that $C_i = C$ and $C_j = D$ and $i \neq j$ because $C \neq D$). Now, if $i > j$, then $C \sqsubseteq C'$, $C \sqsubseteq C''$ and $C' \oplus C''$ (Definition 2.2(3) on page 12). Assume, w.l.o.g., that $C_{i-1} = C'$ this means that $C''^{\hat{I}} \subseteq C^{\hat{I}}$ and $C''^{\hat{I}} \nsubseteq C'^{\hat{I}}$, but since $i > j$, this means that $C''^{\hat{I}} \nsubseteq D^{\hat{I}}$, thus $C^{\hat{I}} \nsubseteq D^{\hat{I}}$ which contradicts the hypothesis. Hence, $j > i$ then $D \sqsubseteq D'$, $D \sqsubseteq D''$ and $D' \oplus D''$ (Definition 2.2(3) on page 12). Assume, again w.l.o.g., that $D_{i-1} = D'$ this means that $D''^{\hat{I}} \subseteq D^{\hat{I}}$ and $D''^{\hat{I}} \nsubseteq D'^{\hat{I}}$, but since $j > i$, this means that $D''^{\hat{I}} \nsubseteq C^{\hat{I}}$, thus $D^{\hat{I}} \nsubseteq C^{\hat{I}}$. □

Proof of Proposition 3.3.
We assume that $\tau(s) \models K_a\phi$ for an ARG state s that is locally consistent and have to show that $\mathcal{O}_a \models \phi$.

[$\phi = C(o)$] Consider that $\tau(s) \models K_a(C(o))$. There exists a unique D such that $D(o) \in \mathcal{O}_a$ (Definition 2.2(4) on page 12) and hence $\mathcal{O}_a \models D(o)$. If $C = D$, then the statement is proven. Otherwise one of the following holds: (i) $\mathcal{O}_a \models C \sqsubseteq D$, (ii) $\mathcal{O}_a \models D \sqsubseteq C$ or (iii) $\mathcal{O}_a \models C \oplus D$ ([19]). Case (i) is impossible because, since C and D are different, D cannot be the (unique) most specific class of object o. In case (iii), by the forward direction, $\tau(s) \models K_a(C \oplus D)$ and $\tau(s) \models K_a(D(o))$. Hence, in every pointed model $\langle \mathcal{M}, w \rangle$ of $\tau(s)$ and at each world v accessible for a from w, $o^{I_v} \in C^{I_v}$, $o^{I_v} \in D^{I_v}$, and $C^{I_v} \cap D^{I_v} = \varnothing$. There can exist no such interpretation.

Thus case (ii) holds, which means that $\mathcal{O}_a \models D \sqsubseteq C$, which together with $\mathcal{O}_a \models D(o)$ entails $\mathcal{O}_a \models C(o)$.

[$\phi = CRD$] Next, consider that ϕ states the relation between two classes C and D, where $\tau(s) \models K_a(\phi)$. If $C = D$, it is clear that $\tau(s) \models K_a(C \sqsubseteq D)$, $\tau(s) \models K_a(D \sqsubseteq C)$ and $\tau(s) \not\models K_a(C \oplus D)$ (because in a model $\langle \mathcal{M}, w \rangle$ of $\tau(s)$ and in any world v accessible for a from w, the interpretation of the class will be the same and non-empty) and for the same reasons, it is clear that $O_a \models C \sqsubseteq D$, $O_a \models D \sqsubseteq C$ and $O_a \not\models C \oplus D$.

Otherwise ($C \neq D$), we know that ([19]) $\mathcal{O}_a \models \psi$ for either (i) $\psi = C \sqsubseteq D$, (ii) $\psi = D \sqsubseteq C$ or (iii) $\psi = C \oplus D$, and (from the forward direction) $\tau(s) \models K_a(\psi)$. In addition, the interpretation of C and D cannot be empty ([19]) so there exists $o, o' \in \mathcal{D}$ such that $\mathcal{O}_a \models C(o)$ and $\mathcal{O}_a \models D(o')$ and thus (by the forward direction) $\tau(s) \models K_a(C(o)) \wedge K_a(D(o'))$. Hence, in each model $\langle \mathcal{M}, w \rangle$ of $\tau(s)$ and in each world v accessible for a from w, $\mathcal{M}, v \models \phi$, $\mathcal{M}, v \models \psi$, and $\mathcal{M}, v \models C(o) \wedge D(o')$.

- If $\phi = C \oplus D$, then the interpretation I_v cannot be such that $\psi = C \sqsubseteq D$ or $\psi = D \sqsubseteq C$ because otherwise $C^{I_v} \cap D^{I_v}$ would contain either o^{I_v} or o'^{I_v}. Hence, it must be that $\psi = C \oplus D$ and thus $\mathcal{O}_a \models C \oplus D$.

- If $\phi = C \sqsubseteq D$, then like before no interpretation can accommodate $\psi = C \oplus D$. To exclude $\psi = D \sqsubseteq C$, we show that there exists a model $\langle \mathcal{M}, w \rangle$ of $\tau(s)$ such that $\mathcal{M}, w \not\models K_a(D \sqsubseteq C)$. Because then, by the forward direction (Proposition 3.1), also $\mathcal{O}_a \not\models D \sqsubseteq C$ and $\mathcal{O}_a \models C \sqsubseteq D$ ([19]). Since s is locally consistent, we can construct a DEOL model $\langle \mathcal{M}^s, w^s \rangle$ of s (Definition 3.6 on page 55) such that $\mathcal{M}^s, w^s \models \tau(s)$ and $\mathcal{M}^s, w^s \models K_a(C \sqsubseteq D)$, i.e. $C^{I_w} \subseteq D^{I_w}$ for each $w \in \{w^s, w_a\}$ – in particular w^s. Because $C^{I_{w^s}} = C^{I_a}$ and $D^{I_{w^s}} = D^{I_a}$ it follows that $C^{I_a} \subseteq D^{I_a}$. Hence, by Lemma A.1, $D^{I_a} \not\subseteq C^{I_a}$ and thus $D^{I_{w^s}} \not\subseteq C^{I_{w^s}}$. In other words, $\mathcal{M}^s, w^s \not\models D \sqsubseteq C$, $\mathcal{M}^s, w^s \not\models K_a(D \sqsubseteq C)$ and thus $\tau(s) \not\models K_a(D \sqsubseteq C)$. Hence $\mathcal{O}_a \models C \sqsubseteq D$.

- Finally, if $\phi = D \sqsubseteq C$, exchanging the symbols C and D in the previous case completes the proof. □

Bibliography

[1] Aberer, K., Cudré-Mauroux, P., Hauswirth, M.: Start making sense: The chatty web approach for global semantic agreements. Journal of Web Semantics **1**(1), 89–114 (2003). DOI 10.1016/j.websem.2003.09.001

[2] Ågotnes, T., Alechina, N.: A logic for reasoning about knowledge of unawareness. Journal of Logic, Language and Information **23**(2), 197–217 (2014)

[3] Alchourrón, C.E., Gärdenfors, P., Makinson, D.: On the logic of theory change: Partial meet contraction and revision functions. The Journal of Symbolic Logic **50**(2), 510–530 (1985)

[4] Atencia, M., Schorlemmer, W.M.: An interaction-based approach to semantic alignment. Journal of Web Semantics **12**, 131–147 (2012). DOI 10.1016/j.websem.2011.12.001

[5] Axelrod, R.: The dissemination of culture: A model with local convergence and global polarization. Journal of conflict resolution **41**(2), 203–226 (1997)

[6] Baader, F., Calvanese, D., McGuinness, D.L., Nardi, D., Patel-Schneider, P.F. (eds.): The Description Logic Handbook: Theory, Implementation, and Applications. Cambridge University Press (2003)

[7] Baltag, A., Moss, L.S., Solecki, S.: The logic of public announcements and common knowledge and private suspicions. In: Proceedings of the 7th Conference on Theoretical Aspects of Rationality and Knowledge, Evanston, USA, 1998, pp. 43–56 (1998)

[8] Baltag, A., Smets, S.: Dynamic belief revision over multi-agent plausibility models. In: Proceedings of the 2016 conference on Logic and the Foundations of Game and Decision Theory, vol. 6, pp. 11–24. University of Liverpool (2006)

[9] Baltag, A., Smets, S.: Group belief dynamics under iterated revision: fixed points and cycles of joint upgrades. In: Proceedings of the 12th Conference on Theoretical Aspects of Rationality and Knowledge, pp. 41–50. ACM (2009)

[10] Baltag, A., Smets, S., et al.: The logic of conditional doxastic actions. Texts in Logic and Games, Special Issue on New Perspectives on Games and Interaction **4**, 9–31 (2008)

[11] Baral, C., Zhang, Y.: Knowledge updates: Semantics and complexity issues. Artificial Intelligence **164**(1-2), 209–243 (2005)

[12] van Benthem, J.: Dynamic logic for belief revision. Journal of Applied Non-Classical Logics **17**(2), 129–155 (2007). DOI 10.3166/jancl.17.129-155

[13] van Benthem, J.: Logical dynamics of information and interaction. Cambridge University Press (2011)

[14] van Benthem, J., van Eijck, J., Kooi, B.: Logics of communication and change. Information and Computation **204**(11), 1620–1662 (2006)

[15] van Benthem, J., Velázquez-Quesada, F.R.: The dynamics of awareness. Synthese **177**(1), 5–27 (2010)

[16] van den Berg, L.: Forgetting agent awareness: a partial semantics approach. In: 4th Women in Logic workshop, pp. 18–21 (2020)

[17] van den Berg, L., Atencia, M., Euzenat, J.: Agent ontology alignment repair through dynamic epistemic logic. In: Proceedings of the 19th International Conference on Autonomous Agents and Multi-Agent Systems, pp. 1422–1430 (2020)

[18] van den Berg, L., Atencia, M., Euzenat, J.: Unawareness in multi-agent systems with partial valuations. In: 10th AAMAS workshop on Logical Aspects of Multi-Agent Systems (2020)

[19] van den Berg, L., Atencia, M., Euzenat, J.: A logical model for the ontology alignment repair game. Autonomous Agents and Multi-Agent Systems **35**(2), 1–34 (2021)

[20] van den Berg, L., Atencia, M., Euzenat, J.: Raising awareness without disclosing truth (submitted)

[21] van den Berg, L., Gattinger, M.: Dealing with unreliable agents in dynamic gossip. In: International Workshop on Dynamic Logic, pp. 51–67. Springer (2020)

[22] Blackburn, P., van Benthem, J., Wolter, F. (eds.): Handbook of Modal Logic, *Studies in logic and practical reasoning*, vol. 3. North-Holland (2007). URL `https://www.sciencedirect.com/bookseries/studies-in-logic-and-practical-reasoning/vol/3/suppl/C`

[23] Blackburn, P., De Rijke, M., Venema, Y.: Modal Logic. Cambridge Tracts in Theoretical Computer Science. Cambridge University Press (2001). DOI 10.1017/CBO9781107050884

[24] Blanché, R.: Reviewed work: Time and modality by A. N. Prior. Revue Philosophique de la France et de l'Étranger **148**, 114–115 (1958)

[25] Blanché, R.: Reviewed work: Some problems of self-reference in John Buridan by A. N. Prior. Revue Philosophique de la France et de l'Étranger **157**, 417–418 (1967)

[26] Borgida, A., Imielinski, T.: Decision making in commitees-a framework for dealing with inconsistency and non-monotonicity. In: Proceedings of the Non-Monotonic Reasoning Workshop, pp. 21–32 (1984)

[27] Brouwer, L.E.J.: Over de grondslagen der wiskunde. Maas & van Suchtelen (1907)

[28] Bylander, T., Chandrasekaran, B.: Generic tasks for knowledge-based reasoning: the "right" level of abstraction for knowledge acquisition. International Journal of Man-Machine Studies **26**(2), 231–243 (1987)

[29] Cavalli-Sforza, L.L., Feldman, M.W.: Cultural transmission and evolution: A quantitative approach. Princeton University Press (1981)

[30] Chang, K.: A math problem from singapore goes viral: When is cheryl's birthday. The New York Times **14** (2015)

[31] Chocron, P., Schorlemmer, M.: Attuning ontology alignments to semantically heterogeneous multi-agent interactions. In: Proceedings of the 22nd European Conference on Artificial Intelligence, The Hague, The Netherlands, pp. 871–879 (2016). DOI 10.3233/978-1-61499-672-9-871

[32] Chocron, P., Schorlemmer, M.: Vocabulary alignment in openly specified interactions. In: Proceedings of the 16th Conference on Autonomous

Agents and Multi-Agent Systems, São Paulo, Brazil, 2017, pp. 1064–1072 (2017). URL http://dl.acm.org/citation.cfm?id=3091275

[33] Chocron, P.D., Schorlemmer, M.: Vocabulary alignment in openly specified interactions. Journal of Artificial Intelligence Research **68**, 69–107 (2020)

[34] Delgrande, J., Lang, J., Schaub, T.: Belief change based on global minimisation. In: 20th International Joint Conference on Artificial Intelligence, pp. 2462–2467 (2007)

[35] Derex, M., Beugin, M.P., Godelle, B., Raymond, M.: Experimental evidence for the influence of group size on cultural complexity. Nature **503**(7476), 389–391 (2013)

[36] van Diggelen, J., Beun, R.J., Dignum, F., van Eijk, R., Meyer, J.J.: Ontology negotiation in heterogeneous multi-agent systems: The ANEMONE system. Applied Ontology **2**(3–4), 267–303 (2007)

[37] van Ditmarsch, H., French, T.: Awareness and forgetting of facts and agents. In: 2009 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology, vol. 3, pp. 478–483. IEEE (2009)

[38] van Ditmarsch, H., French, T.: Becoming aware of propositional variables. In: Indian Conference on Logic and Its Applications, pp. 204–218. Springer (2011)

[39] van Ditmarsch, H., French, T., Velázquez-Quesada, F.R.: Action models for knowledge and awareness. In: Proceedings of the 2012 International Conference on Autonomous Agents and Multi-Agent Systems, pp. 1091–1098 (2012)

[40] van Ditmarsch, H., French, T., Velázquez-Quesada, F.R., Wáng, Y.N.: Knowledge, awareness, and bisimulation. arXiv preprint arXiv:1310.6410 (2013)

[41] van Ditmarsch, H., Herzig, A., Lang, J., Marquis, P.: Introspective forgetting. Synthese **169**(2), 405–423 (2009)

[42] van Ditmarsch, H., van der Hoek, W., Kooi, B.: Dynamic Epistemic Logic, vol. 337. Springer Science & Business Media (2007)

[43] Doyle, J.: A society of mind. Department of Computer Science, Carnegie-Mellon University (1983)

[44] Eiter, T., Ianni, G., Schindlauer, R., Tompits, H., Wang, K.: Forgetting in managing rules and ontologies. In: 2006 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2006 Main Conference Proceedings)(WI'06), pp. 411–419. IEEE (2006)

[45] Euzenat, J.: Semantic precision and recall for ontology alignment evaluation. In: 20th International Joint Conference on Artificial Intelligence, pp. 348–353 (2007)

[46] Euzenat, J.: First experiments in cultural alignment repair (extended version). In: European Semantic Web Conference, pp. 115–130. Springer (2014)

[47] Euzenat, J.: Revision in networks of ontologies. Artificial Intelligence **228**, 195–216 (2015)

[48] Euzenat, J.: Crafting ontology alignments from scratch through agent communication. In: International Conference on Principles and Practice of Multi-Agent Systems, pp. 245–262. Springer (2017)

[49] Euzenat, J.: Interaction-based ontology alignment repair with expansion and relaxation. In: 26th International Joint Conference on Artificial Intelligence, pp. 185–191. AAAI Press (2017)

[50] Euzenat, J., Shvaiko, P.: Ontology Matching, Second Edition. Springer (2013)

[51] Fagin, R., Halpern, J.Y.: Belief, awareness, and limited reasoning. Artificial Intelligence **34**(1), 39–76 (1987)

[52] Fagin, R., Halpern, J.Y., Vardi, M.Y., Moses, Y.: Reasoning about knowledge (1995)

[53] Fensel, D.: Ontologies: a Silver Bullet for Knowledge Management and Electronic Commerce. Springer-Verlag Berlin Heidelberg (2004)

[54] Gärdenfors, P., Rott, H., Gabbay, D., Hogger, C., Robinson, J.: Belief revision. Computational Complexity **63**(6), 35–132 (1995)

[55] Gerbrandy, J.D.: Bisimulations on planet Kripke. Institute for Logic, Language and Computation, University of Amsterdam (1999)

[56] Halpern, J.Y.: Alternative semantics for unawareness. Games and Economic Behavior **37**(2), 321–339 (2001)

[57] Halpern, J.Y., Rêgo, L.C.: Reasoning about knowledge of unawareness. Games and Economic Behavior **67**(2), 503–525 (2009)

[58] Hansen, J.U.: Modeling truly dynamic epistemic scenarios in a partial version of DEL. The Logica Yearbook 2013 pp. 63–75 (2014)

[59] Heifetz, A., Meier, M., Schipper, B.C.: Interactive unawareness. Journal of economic theory **130**(1), 78–94 (2006)

[60] Hill, B.: Awareness dynamics. Journal of Philosophical Logic **39**(2), 113–137 (2010)

[61] Hintikka, J.: Knowledge and belief: An introduction to the logic of the two notions. Studia Logica **16** (1962)

[62] van der Hoek, W., Jaspars, J., Thijsse, E.: Honesty in partial logic. Studia Logica **56**(3), 323–360 (1996)

[63] van der Hoek, W., Wooldridge, M.: Multi-agent systems. Foundations of Artificial Intelligence **3**, 887–928 (2008)

[64] Jaspars, J., Thijsse, E.: Fundamentals of partial modal logic. Studies in Logic Language and Information (1996)

[65] Jiménez-Ruiz, E., Meilicke, C., Grau, B.C., Horrocks, I.: Evaluating mapping repair systems with large biomedical ontologies. Description Logics **13**, 246–257 (2013)

[66] Jiménez-Ruiz, E., Payne, T., Solimando, A., Tamma, V.: Limiting logical violations in ontology alignnment through negotiation. In: Principles of Knowledge Representation and Reasoning: Proceedings of the Fifteenth International Conference, Cape Town, South Africa, pp. 217–226 (2016). URL `http://www.aaai.org/ocs/index.php/KR/KR16/paper/view/12893`

[67] Kleene, S.C.: Introduction to metamathematics. North-Holland, Amsterdam (NL) (1952)

[68] Konev, B., Walther, D., Wolter, F.: Forgetting and uniform interpolation in large-scale description logic terminologies. In: Twenty-First International Joint Conference on Artificial Intelligence (2009)

[69] Laera, L., Blacoe, I., Tamma, V., Payne, T., Euzenat, J., Bench-Capon, T.: Argumentation over ontology correspondences in MAS. In: Proceedings of the 6th international joint conference on Autonomous Agents and Multi-Agent Systems, pp. 1–8 (2007)

[70] Lin, F., Reiter, R.: Forget it. In: Working Notes of AAAI Fall Symposium on Relevance, pp. 154–159 (1994)

[71] Liu, F., Seligman, J., Girard, P.: Logical dynamics of belief change in the community. Synthese **191**(11), 2403–2431 (2014)

[72] Lutz, C., Wolter, F.: Foundations for uniform interpolation and forgetting in expressive description logics. In: Twenty-Second International Joint Conference on Artificial Intelligence (2011)

[73] Meilicke, C.: Alignment incoherence in ontology matching. Ph.D. thesis, University of Mannheim (2011). URL `https://ub-madoc.bib.uni-mannheim.de/29351`

[74] Meilicke, C., Stuckenschmidt, H.: Incoherence as a basis for measuring the quality of ontology mappings. In: Proc. 3rd International Workshop on Ontology Matching (OM) co-located with ISWC, pp. 1–12 (2008)

[75] Mesoudi, A., Whiten, A., Laland, K.N.: Towards a unified science of cultural evolution. Behavioral and Brain Sciences **29**(4), 329–347 (2006)

[76] Meyer, J.J.C., Van Der Hoek, W.: Epistemic logic for AI and computer science. 41. Cambridge University Press (2004)

[77] Modica, S., Rustichini, A.: Awareness and partitional informational structures. In: Epistemic Logic and the Theory of Games and Decisions, pp. 151–168. Springer (1997)

[78] Moore, R.C., Hendrix, G.G.: Computational models of belief and the semantics of belief sentences. In: Processes, Beliefs, and Questions, pp. 107–127. Springer (1982)

[79] Mossakowski, T., Kutz, O., Codescu, M., Lange, C.: The distributed ontology, modelling and specification language–dol. In: Workshop on Modular Ontologies 2013, p. 1 (2013)

[80] Payne, T., Tamma, V.: Negotiating over ontological correspondences with asymmetric and incomplete knowledge. In: Proceedings of the 2014 international conference on Autonomous Agents and Multi-Agent Systems, pp. 517–524 (2014)

[81] Plaza, J.: Logic of public communications. In: Z.W. Ras (ed.) Proceedings of the 4th International Symposium on Methodologies for Intelligent Systems, pp. 201–216. North-Holland (1989)

[82] Plaza, J.: Logics of public communications. Synthese **158**(2), 165–179 (2007). DOI 10.1007/s11229-007-9168-7

[83] Rendsvig, R., Symons, J.: Epistemic Logic. In: E.N. Zalta (ed.) The Stanford Encyclopedia of Philosophy, Summer 2021 edn. Metaphysics Research Lab, Stanford University (2021)

[84] Richerson, P.J., Boyd, R.: Not by genes alone: How culture transformed human evolution. University of Chicago press (2008)

[85] Rott, H.: Conditionals and theory change: Revisions, expansions, and additions. Synthese **81**(1), 91–113 (1989)

[86] Trojahn dos Santos, C., Euzenat, J., Tamma, V., Payne, T.: Argumentation for reconciling agent ontologies. In: A. Elçi, M. Koné, M. Orgun (eds.) Semantic Agent Systems, chap. 5, pp. 89–111. Springer, New-York (NY US) (2011)

[87] Santos, E., Faria, D., Pesquita, C., Couto, F.M.: Ontology alignment repair through modularization and confidence-based heuristics. PloS One **10**(12), e0144807 (2015)

[88] Schwind, N., Inoue, K., Bourgne, G., Konieczny, S., Marquis, P.: Belief revision games. In: Proceedings of the AAAI Conference on Artificial Intelligence, pp. 1590–1596 (2015)

[89] Shvaiko, P., Euzenat, J.: Ontology matching: state of the art and future challenges. IEEE Transactions on Knowledge and Data Engineering **25**(1), 158–176 (2013)

[90] Smith, J.M.: The problems of biology. Oxford University Press (1986)

[91] Steels, L. (ed.): Experiments in cultural language evolution, vol. 3. John Benjamins Publishing (2012)

[92] Su, K., Lv, G., Zhang, Y.: Reasoning about knowledge by variable forgetting. KR **4**, 576–586 (2004)

[93] Tamma, V., Cranefield, S., Finin, T.W., Willmott, S.: Ontologies for agents: Theory and experiences. Springer Science & Business Media (2005)

[94] Thijsse, E.: Partial logic and knowledge representation. Ph.D. thesis, Tilburg University (1992). Pagination: VIII, 253

[95] Veltman, F.: Defaults in update semantics. Journal of philosophical logic **25**(3), 221–261 (1996)

[96] Wang, Y., Cao, Q.: On axiomatizations of public announcement logic. Synthese **190**(1), 103–134 (2013)

[97] Wang, Z., Wang, K., Topor, R., Pan, J.Z., Antoniou, G.: Uniform interpolation for $\mathcal{ALC}$ revisited. In: Australasian Joint Conference on Artificial Intelligence, pp. 528–537. Springer (2009)

List of Symbols

$a, b, c, \ldots$	agents
$i, j, \ldots$	agent variables
$\mathcal{A}$	set of agents
$p, q, r, \ldots$	atomic propositions
P	set of propositions
$\phi, \psi, \ldots$	formulas
R	relations
R^*	transitive closure of R
$Max_R(S)$	set of maximal elements of S with respect to R
$\neg$	negation
$\wedge, \vee, \rightarrow, \leftrightarrow$	logical connectives (conjunction, disjunction, implication, bi-implication)
$\varnothing$	empty set
$\mathcal{P}(S)$	powerset of S
$Dom(f)$	domain of f
$\forall$	universal quantifier
$\exists$	existential quantifier
$\exists!$	uniqueness quantifier
$\models$	semantic consequence relation for verification
$=\!\!\vert$	semantic consequence relation for falsification
$C, D, \ldots, \top$	classes
$\mathcal{C}$	set of classes
$o, o', \ldots$	objects
$\mathcal{D}$	set of objects
$msc(C)$	the most specific superclass of C
$msc(o)$	the most specific class of object o
$mscc(C, o)$	the most specific superclass of C containing o
$mgcx(C, o)$	the most general subclasses of C not containing o
CRD	relation variable between classes C and D
$\sqsubseteq, \sqsupseteq, \oplus, \between$	class relations (subsumption, reversed subsumption,

	disjointness, overlap)
$\in$	membership relation
$C(o)$	membership abbreviation for $o \in C$
I	interpretation function
C^I	interpretation of class C under interpretation I
o^I	interpretation of object o under interpretation I
Δ	domain of interpretation
$\langle \mathcal{C}, \mathcal{D} \rangle$	ontology signature
$\mathcal{O}$	ontology
$\mathcal{O}_i$	ontology of agent i
A_{ij}	alignment between agents i and j
c	correspondence variable
s	ARG state
$\alpha[c, o], \alpha$	adaptation operator (applied to correspondence c with object o)
`delete`	delete adaptation operator
`add`	add adaptation operator
`addjoin`	addjoin adaptation operator
`refine`	refine adaptation operator
`refadd`	refadd adaptation operator
$\mathcal{L}_{DEL}$	syntax of Dynamic Epistemic Logic (DEL)
$\mathcal{L}_{EAL}$	syntax of Epistemic Action Logic (EAL)
$\mathcal{L}_{DEOL}$	syntax of Dynamic Epistemic Ontology Logic (DEOL)
$\mathcal{L}_{ParDEL}$	syntax of Partial Dynamic Epistemic Logic (ParDEL)
$\mathcal{L}_{ParDEL+}$	syntax of Partial Dynamic Epistemic Logic with raising awareness (ParDEL+)
$\mathcal{L}_{ParEAL}$	syntax of Partial Epistemic Action Logic (ParEAL)
$\mathcal{L}_{ParDEL\odot}$	syntax of Partial Dynamic Epistemic Logic with raising awareness and forgetting (ParDEL$\odot$)
$\mathcal{L}_{ParDEOL}$	syntax of Partial Dynamic Epistemic Ontology Logic (ParDEOL)
$\mathfrak{F}$	Kripke frame
$\mathcal{M}$	Kripke model
$w, v, \ldots$	worlds
W	set of worlds
$\langle \mathcal{M}, w \rangle$	pointed Kripke model
$\leqslant_i$	plausibility relation for agent i
R_i	accessibility relation for agent i
$\sim_i$	epistemic relation for agent i
$\rightarrow_i$	doxastic relation for agent i

$\lvert w\rvert_a$	information (or accessible) cell of agent a at w
$\lvert\lvert w\rvert\rvert_a$	aware cell of agent a at w
V	valuation function
$\lvert\lvert\phi\rvert\rvert_{\mathcal{M}}$	set of worlds in $\mathcal{M}$ making ϕ true
$\dagger\phi$	modal operator variable
$\mathcal{M}^{\dagger\phi}$	modal operator $\dagger\phi$ applied to the model $\mathcal{M}$
$!\phi$	public announcement operator
$!_G\phi$	private announcement operator
$\Uparrow\phi$	radical upgrade
$\uparrow\phi$	conservative upgrade
$\odot p$	raising/forgetting operator variable
$\mathcal{M}^{\odot p}$	raising/forgetting operator $\odot p$ applied to the model $\mathcal{M}$
$+p$	raising (propositional) awareness
$+_G p$	private raising (propositional) awareness
$-p$	forgetting (propositional) awareness
$-_G p$	private forgetting (propositional) awareness
$\ominus p$	forgetting (propositional) truth
$\ominus_G p$	private forgetting (propositional) truth
$\odot C$	raising/forgetting class operator variable
$\mathcal{M}^{\odot C}$	raising/forgetting class operator $\odot C$ applied to the model $\mathcal{M}$
$+C$	raising class awareness
$+_G C$	private raising class awareness
$-C$	forgetting class awareness
$-_G C$	private forgetting class awareness
$\ominus C$	forgetting class truth
$\ominus_G C$	private forgetting class truth
$e, e', \ldots$	events
E	set of event
$\mathcal{E}$	event model
$\mathcal{M} \otimes \mathcal{E}$	product update
τ	translation of ARG states in DEOL
δ	translation of adaptation operators in DEOL
δ^+	translation of adaptation operators in ParDEOL

Résumé

Pour raisonner et parler du monde, les agents peuvent utiliser leurs propres vocabulaires distincts, structurés en représentations de connaissances, également appelées ontologies. Afin de communiquer, ils utilisent des alignements : des traductions entre les termes de leurs ontologies. Cependant, les ontologies peuvent changer, nécessitant que leurs alignements évoluent en conséquence. L'évolution culturelle et expérimentale de la connaissancee offre un cadre pour étudier les mécanismes de l'évolution de leurs connaissances. Il a été appliqué à l'évolution des alignements dans le jeu de réparation d'alignement (ARG).

Des expériences ont montré que, grâce à ARG, les agents améliorent leurs alignements et parviennent à une communication réussie. Pourtant, ces expériences ne sont pas suffisantes pour établir les propriétés formelles de l'évolution des connaissances culturelles.

Cette thèse jette un pont entre l'évolution culturelle de la connaissance et une modèle théorique de l'évolution culturelle de la connaissance en logique. Ceci est réalisé en introduisant la Logique épistémique dynamique d'ontologies (DEOL) et en définissant une traduction fidèle de ARG en DEOL qui (a) encode les ontologies, (b) fait correspondre les ontologies et les alignements des agents aux connaissances et aux croyances, et (c) capture les opérateurs d'adaptation à travers les annonces et les mises à jour conservatrices. Ce modèle montre que tous les opérateurs d'adaptation sauf un sont corrects, qu'ils sont incomplets et que certains sont partiellement redondants.

Trois différences entre les agents ARG et leur modèle logique expliquent ces résultats, conduisant à un modèle de conscience indépendant basé sur des évaluations partielles et des relations faiblement réflexives. Un modèle alternatif d'ARG est alors défini sous lequel les propriétés formelles sont réexaminées, montrant que ce modèle est plus proche du jeu original. Il s'agit d'un premier pas vers la définition d'un modèle théorique d'évolution des connaissances culturelles.

Abstract

To reason and talk about the world, agents may use their own distinct vocabularies, structured into knowledge representations, also called ontologies. In order to communicate, they use alignments: translations between terms of their ontologies. However, ontologies may change, requiring their alignments to evolve accordingly. Experimental cultural evolution offers a framework to study the mechanisms of their knowledge evolution. It has been applied to the evolution of alignments in the Alignment Repair Game (ARG).

Experiments have shown that, through ARG, agents improve their alignments and reach successful communication. Yet, these experiments are not sufficient to understand the formal properties of cultural knowledge evolution.

This thesis bridges experimental cultural knowledge evolution with a theoretical model of cultural knowledge evolution in logic. This is achieved by introducing Dynamic Epistemic Ontology Logic and defining a faithful translation of ARG in DEOL that (a) encodes the ontologies, (b) maps agents' ontologies and alignments to knowledge and beliefs, and (c) captures the adaptation operators through announcements and conservative upgrades. This model shows that all but one adaptation operator are correct, they are incomplete and some are partially redundant.

Three differences between the ARG agents and their logical model explain these results, leading to an independent model of awareness based on partial valuations and weakly reflexive relations. An alternative model of ARG is then defined under which the formal properties are re-examined, showing that this model is closer to the original game. This is a first step towards defining a theoretical model of cultural knowledge evolution.