A logical model for the ontology alignment repair game

Line van den Berg, Manuel Atencia, Jérôme Euzenat

Autonomous Agents and Multi-Agent Systems **35**:32, 2021
DOI:10.1007/s10458-021-09508-8

A Logical Model for the Ontology Alignment Repair Game

Line van den Berg · Manuel Atencia · Jérôme Euzenat

Received: 2020-10-05 / Accepted: 2021-05-20 / Published: 2021-06-26

Abstract Ontology alignments enable agents to communicate while preserving heterogeneity in their knowledge. Alignments may not be provided as input and should be able to evolve when communication fails or when new information contradicting the alignment is acquired. The Alignment Repair Game (ARG) has been proposed for agents to simultaneously communicate and repair their alignments through adaptation operators when communication failures occur. ARG has been evaluated experimentally and the experiments showed that agents converge towards successful communication and improve their alignments. However, whether the adaptation operators are formally correct, complete or redundant could not be established by experiments. We introduce a logical model, Dynamic Epistemic Ontology Logic (DEOL), that enables us to answer these questions. This framework allows us (1) to express the ontologies and alignments used via a faithful translation from ARG to DEOL, (2) to model the ARG adaptation operators as dynamic modalities and (3) to formally define and establish the correctness, partial redundancy and incompleteness of the adaptation operators in ARG.

Keywords Ontology alignment · Alignment repair · Multi-agent systems · Agent communication · Dynamic Epistemic Logic

1 Introduction

Agents may use ontologies or other logical representations of their knowledge of the world they live in. In autonomous agent systems this may lead them to develop

L. van den Berg
Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LIG, F-38000 Grenoble France E-mail: line.van-den-berg@inria.fr

M. Atencia
Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LIG, F-38000 Grenoble France E-mail: manuel.atencia@inria.fr

J. Euzenat
Univ. Grenoble Alpes, Inria, CNRS, Grenoble INP, LIG, F-38000 Grenoble France E-mail: jerome.euzenat@inria.fr

different ontologies. This heterogeneity causes a problem when agents try to communicate: how do agents understand each other if they express their knowledge in different ways?

Ontology matching has been developed to allow agents with different knowledge representations to communicate [23]. Ontology matching algorithms aim to find relationships holding across entities of two ontologies, the *ontology alignment*. Yet, ontology matching algorithms may output only partially correct or incomplete alignments, and there may be situations in which agents constantly evolve their ontologies requiring alignments to evolve accordingly. This means that alignments have to evolve. One way for agents to achieve this is to adapt the alignments when experience shows them that one of them is not correct.

These types of scenarios are of widespread application. For instance, one may consider the situation where agents are taking care of a warehouse that continuously has to take into account the references of new suppliers and to reclassify them in function of different characteristics. Similarly, one may consider social care robots that independently generalise from what they observe and have to coordinate with different such robots. Of course, it is possible to shut down the warehouse and to stop social care waiting for the re-engineering of a common ontology. But this would not be very reactive, nor very open.

The Alignment Repair Game (ARG) [11,13], inspired by cultural language evolution [24], is a model of such situations, used for performing simulations. It allows agents to communicate and, in parallel, evolve their alignments through local corrective actions whenever their communication fails. In ARG, the agents are adaptive: they blindly apply adaptation operators that specify how to modify the failing alignment.

ARG is used to assess the effect of different adaptation operators on agents' knowledge. It was evaluated experimentally and the experiments showed that the agents converge towards successful communication and improve their alignments [13]. However, experiments do not assess formal properties of operators, i.e. whether they are logically *correct*, *complete* or *redundant*.

In this paper we introduce a logical model based on Dynamic Epistemic Logic (DEL) [34] called Dynamic Epistemic Ontology Logic (DEOL) to answer this question. In particular, DEOL enables us to express the ontologies and alignments via a faithful translation and to model the ARG adaptation operators as dynamic modalities. This allows us to formally define correctness, completeness and redundancy for ARG and to use these definitions to prove that all but the `add` operator are correct, that `delete`, `addjoin` and `refine` are redundant for one agent and that all operators are incomplete.

The contributions of this paper are thus to (1) provide a logical framework suitable to model the alignment repair game, (2) formally define the formal properties of the adaptation operators: correctness, redundancy and completeness, (3) prove whether these properties hold for `delete`, `add`, `addjoin`, `refine` and `refadd`, and finally (4) theoretically compare adaptive agents and logical agents.

We restrict the logical modelling to ARG, however, the techniques presented here could be extended to establish formal properties of other games that are designed for agents to improve alignments through interaction.

This paper is an extension of [5] which provides a precise definition of the game, ontologies and alignments (dropping the adaptation operator `replace` without loss of generality to gain clarity); a complete revised section of related work; a

full definition of Dynamic Epistemic Ontology Logic (DEOL) and a proof of the faithfulness of the translation from ARG to DEOL; and a broader discussion of the significance and perspectives of this work.

In the remainder, we discuss the related work (§2) and provide the preliminaries (§3): the Alignment Repair Game (§3.1-3.2) and Dynamic Epistemic Logic (§3.3). We introduce DEOL (§4), translate ARG states to DEOL models and prove that this translation is consequence-preserving (§5). Then, we define the formal properties of the adaptation operators and use these definitions to prove that all but the `add` operator are correct, that`delete`, `addjoin` and `refine` are redundant for one agent and that all operators are incomplete (§6). We conclude by discussing the results and giving a perspective on future work (§7).

2 Related Work

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. Autonomous agents, developed from different sources or learning autonomously, will naturally have different ontologies. This heterogeneity is a problem, in particular when agents need to communicate about this common environment [25].

Imposing a single universally shared ontology on agents would solve the problem but at the price of autonomy. It also violates the possibility of keeping part of the ontology private. It is thus reasonable to consider that not all ontologies have to be shared.

In consequence, a common method relies on ontology alignments [14] to ensure agent intelligibility when communicating: alignments express relations between concepts and relations that allow them to interpret other agents' messages.

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 [31] matches ontologies on the fly when agents need to communicate and cannot express themselves. Agents will then exchange concept definitions or concept instances. Other approaches use argumentation to accept or reject correspondences from a library of alignments [18, 26].

Such techniques consider the obtained alignment as correct and do not consider modifying or repairing it dynamically. Different techniques have been proposed to evolve alignments: gossiping amongst agents to reach global agreement [1], logical repair to enforce consistency [19, 16, 22], or prevention of logical violations to agents' ontologies via conservativity principles [17]. Some have been integrated with multi-agent systems via specific protocols [1, 20].

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 [2]. This is an ontology matching algorithm embedded in the interaction protocols used by agents to communicate. Alignments are induced depending on repeated successful interactions and failing interactions lead to revision. This proposal was further advanced to repair alignments *through their use* and generalized to less constrained protocols [2, 8, 9, 10].

The Alignment Repair Game (ARG) [11], taking inspiration from experimental cultural language evolution [24] may be considered as belonging to this category of approaches. It lets agents continuously play the game and systematically adapt their alignments when a failure occurs. ARG has been used to experimentally show that adaptive agents converge towards successful communication through local corrective actions [13]. Yet, very little of the formal properties of the ARG agents or adaptation operators were assessed formally. This is the purpose of the current work.

3 Preliminaries

In this section, we first explain what is needed of ontologies and ontology alignments to explain ARG (§3.1). Then we describe the Alignment Repair Game as a way for agents to evolve alignments (§3.2). Finally we give the syntax and semantics of Dynamic Epistemic Logic (DEL) (§3.3). This logic is the basis for the logic introduced later in this paper.

ARG may be described in different ways. We opted for a simple presentation that both strictly fits ARG and aligns well with the DEOL translation. This involved two specific choices: having a unique namespace for objects, instead of identifying objects with properties, and having a full network of monodirectional alignments (ARG alignments).

3.1 Ontologies and alignments

Agents use ontologies to represent their knowledge of the world. An ontology is based on a signature.

Definition 1 (Ontology signature) An *ontology signature*, or *signature*, 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 [3]. Here we restrict ourselves to the very simple ontologies manipulated within 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 1. We will consider several agents, named a, b, c, etc., each having their own ontologies. We will use subscript to identify the associated concepts: $\mathcal{O}_a$, $\mathcal{C}_a$, $mgcx_a(C, o)$, etc.

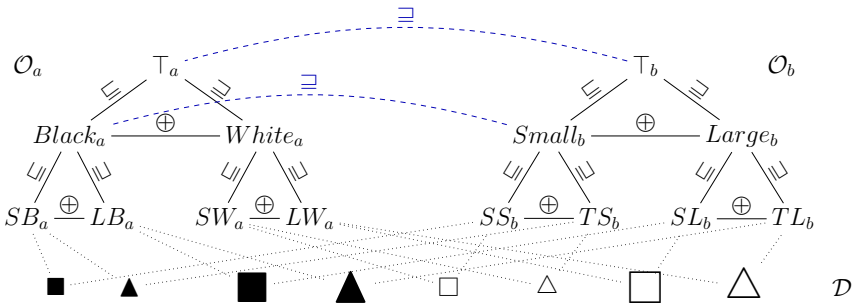

Fig. 1 Two ARG ontologies, $\mathcal{O}_a$ on the left and $\mathcal{O}_b$ on the right, and an alignment A_{ab} (in dashed blue) between them. Membership between objects and classes are rendered by dotted edges. The class names of the leaf classes have an intended use for the agents that use them: SB_a 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.

ARG ontologies are formally defined in Definition 2 on which proofs are based. Such constraints may be obtained in other ways, but the present ones are sufficient for this paper.

Definition 2 (ARG ontology) An *ARG ontology* $\mathcal{O}$ over a signature $\langle \mathcal{C}, \mathcal{D} \rangle$ is a finite set of axioms following the grammar:

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

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:
   – $\exists! \langle C, C' \rangle \in \mathcal{C} \times \mathcal{C}$ with $C \sqsubseteq D \in \mathcal{O}$, $C' \sqsubseteq D \in \mathcal{O}$ and $C \oplus C' \in \mathcal{O}$, or
   – $\nexists C \in \mathcal{C}$ 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 specify that (1) there is a most general class $\top$, (2) each class, apart from $\top$, has a unique class to which it is subsumed, (3) all classes either have a unique pair of subclasses that are disjoint, or there is an object that belongs to it, (4) each object has a unique class to which it belongs, and (5) there is no cycle.

In the following, the word 'ontology' denotes an ARG ontology. Definition 2 constrains the syntax of ARG ontologies. An *interpretation* is provided to give meaning to the elements of the signature.

Definition 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 ($\cdot^I : \mathcal{C} \to 2^\Delta$), such that $\top^I = \Delta$.

Definition 4 (Formula Satisfiability) Let $\mathcal{O}$ be an ontology and let I be an interpretation over its signature, satisfiability ($\models$) is defined as follows:

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

We may use $I \models C \sqsubset D$ whenever $I \models C \sqsubseteq D$ but $I \not\models C \sqsupseteq D$ 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.

Proposition 1 *Every ARG ontology is consistent.*

This is a good reason for ARG agents to never change their ontologies.

We introduce some notation useful for defining ARG games precisely:

Definition 5 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:

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

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 $mgcx(C, o)$.

Example 1 (Running example) Considering the ontologies of Figure 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 $\triangle$.

These concepts are well-defined for ARG ontologies:

Lemma 1 *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.

Because agents use different vocabularies in their ontologies, they face problems when trying to communicate: how do they understand each other if they express their knowledge in different ways? Alignments are tools that allow agents to translate their knowledge with respect to the ontology of other agents, enabling them to communicate with each other [14].

Definition 6 (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 paper, alignments are shared between the involved agents and we consider here 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 [14]. 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}). Hence we define the notion of local entailment that only one of the agents has access to.

Definition 7 (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 $\models_a$) as follows:

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

such that $\cdot^{I'}$ is an extension of $\cdot^I$ to $\mathcal{C}_a \cup \mathcal{C}_b$.

The definition can be rewritten for covering how b interprets A_{ab}.

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 $\mathcal{L}$ is a model of $\mathcal{O}$ an extension of which locally satisfies all the correspondences in all alignments in $\mathcal{L}$. An alignment is *locally consistent* if it has a local model (otherwise locally inconsistent), an ontology is *locally consistent* for a set of alignments $\mathcal{L}$ if it has a local model for $\mathcal{L}$, 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 $A \models_a \gamma$).

Even such a restricted definition 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 $\{\langle C, D, \sqsubseteq\rangle, \langle C', D, \sqsupseteq\rangle\} \subseteq A_{ab}$, 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.

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 [12]. Moreover, the alignments have to follow specific constraints allowing to always find a single correspondence applicable for an object.

Definition 8 (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 is 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 9 (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 $\mathcal{O}_b$.

An ARG state is said *locally consistent* for a set of agents $\mathcal{A}$ if it is locally consistent for each agent in $\mathcal{A}$. We will say that an ARG state is *globally consistent* 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 [12].

3.2 Alignment Repair Game

The Alignment Repair Game (ARG) is a protocol designed for adaptive agents to evolve alignments between their ontologies through their use [11, 13]. 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 10 (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 that will serve as a running example throughout this paper.

Example 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 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 $\mathcal{O}_a \models Black_a(▲)$), for the object △ a failure is reached (because $\mathcal{O}_a \models White_a(△) \wedge White_a \oplus Black_a$). 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 3).

The agent behaviour, 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 8), the agent does not choose the correspondence to apply.

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

Definition 11 (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.

In [11,13] 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`$[\langle C_a, C_b, \sqsupseteq\rangle, o]$, adds 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`$[\langle C_a, C_b, \sqsupseteq\rangle, o]$: the combination of the operators `addjoin`$[\langle C_a, C_b, \sqsupseteq\rangle, o]$ and `refine`$[\langle C_a, C_b, \sqsupseteq\rangle, o]$.

Formally, this amounts to:

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

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. Example 3 illustrates the application of these operators.

Example 3 (Running example) In the situation reached at the end of Example 2 with object $\triangle$, an adaptation operator $\alpha[\langle Black_a, Small_b, \sqsupseteq\rangle, \triangle]$ is applied to the alignment A_{ab}, deleting the initial correspondence and adding 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`: $\langle Black_a, SS_b, \sqsupseteq\rangle$ $(mscc_b(Small_b, \triangle) = SS_b)$
- `refadd`: $\langle \top_a, Small_b, \sqsupseteq\rangle$ and $\langle Black_a, SS_b, \sqsupseteq\rangle$

From the definition, 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 illustrates the effect of the adaptation operators.

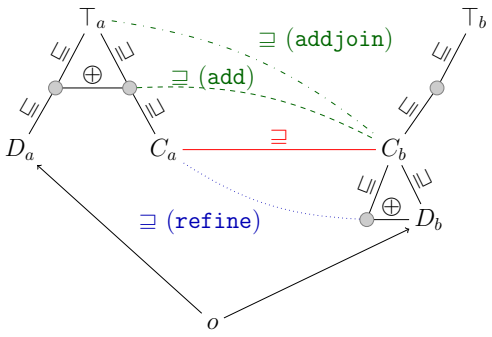

Fig. 2 Schematic diagram of the deleted (red, solid) and added correspondences (green and dashed for `add`, green and dash-dotted for `addjoin`, blue and dotted for `refine`) by the different adaptation operators in ARG.

The alignment repair game modifies the situation from ARG state to ARG state, as expressed by Property 1.

Property 1 (Operators preserve ARG stateness) Given an ARG state s and a failure of correspondence c on object o, then $\alpha[c, o](s)$ is well-defined and is an ARG state, for each operator $\alpha \in \{\texttt{delete}, \texttt{add}, \texttt{addjoin}, \texttt{refine}, \texttt{refadd}\}$.

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

By playing ARG with different operators, they can be compared. In [11,13], the operators are compared experimentally in terms of success rate (ratio of successes over rounds played), semantic precision and recall with respect to the known correct reference alignment (the degree of correctness and completeness of the resulting alignment) and convergence (the number of rounds needed to converge). It was found that all the operators have a relatively high success rate, yet do not reach 100% precision, and that recall and convergence both increases with operators that add new correspondences. The operator `refadd`, followed by `add`, shows the highest semantic recall and `add`, the slowest convergence.

In order to study the formal properties of ARG adaptation operators, we introduce dynamic epistemic logics.

3.3 Dynamic Epistemic Logic

Dynamic Epistemic Logics (DEL) are a family of modal logics describing information flow in multi-agent systems. DEL has been widely used as a formal framework to model agent communication [21,29,34], belief revision [27] and agent interaction [28]. As such, it provides a solid basis to study knowledge and belief evolution of logical agents playing ARG. Here we consider the syntax and semantics introduced by [4].

Definition 12 (Syntax of DEL) The *syntax*, $\mathcal{L}_{DEL}$, of (multi-agent) DEL 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 $p \in \mathcal{P}$ is a proposition, K_a and B_a are the knowledge and belief operators for each agent a and $\dagger\phi$ with $\dagger \in \{!, \Uparrow\}$ the dynamic upgrades.

The connectives $\vee$ and $\rightarrow$, and the duals $\hat{K}_a, \hat{B}_a, \langle\dagger\phi\rangle$ are defined in the usual way: $\phi \vee \psi$ iff $\neg(\neg\phi \wedge \neg\psi)$, $\phi \rightarrow \psi$ iff $\neg\phi \vee \psi$, $\hat{K}_a\phi = \neg K_a\neg\phi$, $\hat{B}_a\phi = \neg B_a\neg\phi$, and $\langle\dagger\phi\rangle = \neg[\dagger\phi]\neg\psi$. DEL models are based on Kripke frames with plausibility relations where the dynamic modalities act as model transformers.

Definition 13 (DEL Model) A *model* of (multi-agent) DEL is a triple $\mathcal{M} = \langle W, (\geqslant_a)_{a\in\mathcal{A}}, V\rangle$ where

- W is a non-empty set of worlds;
- $(\geqslant_a)_{a\in\mathcal{A}} \subseteq W \times W$ are the plausibility relations on W, one for each agent, that are converse well-founded, locally connected preorders;
- and V is a propositional valuation mapping propositions to sets of worlds in which that proposition is true.

The plausibility relation $w \geqslant_a v$ reads as "w is at least as plausible as v for agent a" and the epistemic and doxastic relations are defined on W accordingly:

$$w \sim_a v \text{ iff } w\ (\leqslant_a \cup \geqslant_a)\ v \tag{1}$$

$$w \rightarrow_a v \text{ iff } v \in Max_{\leqslant_a}|w|_a \tag{2}$$

where $|w|_a$ is the *information cell* (or *accessible cell*) of agent a at state w and is defined by:

$$|w|_a = \{v \in W \mid w \sim_a v\} \tag{3}$$

It follows from the properties of $\leqslant_a$ and $\geqslant_a$ that the relations $\sim_a$ are reflexive, transitive and symmetric, and the relations $\rightarrow_a$ are transitive, serial and Euclidean. Therefore they satisfy the usual properties of knowledge and belief, respectively [7, 34].

Satisfiability is considered with respect to a pointed model $\langle \mathcal{M}, w \rangle$ which associates a DEL model $\mathcal{M}$ with a world w.

Definition 14 (Satisfiability for DEL) *Satisfiability* for DEL by a pointed model $\langle \mathcal{M}, w \rangle$ is defined in the following way:

$$\begin{array}{ll}
\mathcal{M}, w \models p & \text{iff } w \in V(p) \\
\mathcal{M}, w \models \phi \wedge \psi & \text{iff } \mathcal{M}, w \models \phi \text{ and } \mathcal{M}, w \models \psi \\
\mathcal{M}, w \models \neg\phi & \text{iff } \mathcal{M}, w \not\models \phi \\
\mathcal{M}, w \models K_a\phi & \text{iff } \forall v \text{ s.t. } w \sim_a v : \mathcal{M}, v \models \phi \\
\mathcal{M}, w \models B_a\phi & \text{iff } \forall v \text{ s.t. } w \rightarrow_a v : \mathcal{M}, v \models \phi \\
\mathcal{M}, w \models [!\phi]\psi & \text{iff } \mathcal{M}^{!\phi}, w \models \psi \\
\mathcal{M}, w \models [\uparrow\phi]\psi & \text{iff } \mathcal{M}^{\uparrow\phi}, w \models \psi
\end{array}$$

where $!\phi$ and $\uparrow\phi$ are model transformers $!\phi : \mathcal{M} \rightarrow \mathcal{M}^{!\phi}$ and $\uparrow\phi : \mathcal{M} \rightarrow \mathcal{M}^{\uparrow\phi}$ with $||\phi||_{\mathcal{M}} = \{w \in W \mid \mathcal{M}, w \models \phi\}$ such that

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

Conservative upgrade ($\uparrow \phi$) Change the plausibility orders so that the best 'ϕ'-worlds become better than all other worlds, while the old ordering on the rest of the worlds remains, i.e. $W^{\uparrow\phi} = W$, $w \geqslant_a^{\uparrow\phi} v$ iff either $w \in Max_{\leqslant_a}(|v|_a \cap ||\phi||_{\mathcal{M}})$ or $w \geqslant_a v$, $V^{\uparrow\phi}(p) = V(p)$.

We also write $\dagger_1\phi; \dagger_2\psi$ for the sequence of upgrades $\dagger_1\phi$ and then $\dagger_2\psi$. The resulting model $\mathcal{M}^{\dagger_1\phi;\dagger_2\psi}$ is equal to $(\mathcal{M}^{\dagger_1\phi})^{\dagger_2\psi}$.

The intuition behind the different upgrades is that the trustworthiness of the information source may vary: it may be considered from an infallible source (announcements), or from a trusted, but not infallible source (conservative upgrades). For this reason, conservative upgrades only change the plausibility of worlds without deleting any alternatives. This also means that an announcement $!\phi$ can only be validly performed on a model $\mathcal{M}$ if ϕ is true there. Since the worlds do not change, it is not necessary to apply the announcements to w when interpreting them.

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

4 Dynamic Epistemic Ontology Logic

Dynamic Epistemic Logic (DEL) is the study of epistemic, or more generally modal, logics that are subjected to model-transforming actions. Modelling the communication taking place in ARG requires a similar approach. Here, we introduce Dynamic Epistemic Ontology Logic (DEOL) that is 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.

Definition 15 (Syntax of DEOL) The *syntax*, $\mathcal{L}_{DEOL}$, of (multi-agent) DEOL is defined in the following way:

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

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

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 $\dagger \in \{!, \Uparrow\}$ 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 in the case for DEL.

The difference with DEL models is that instead of a valuation of propositions, we consider a domain of interpretation Δ representing the objects and an interpretation function I assigning to each world a function interpreting each class as a set of objects of the domain.

Definition 16 (DEOL Model) A *model* of (multi-agent) DEOL is a quadruple $\mathcal{M} = \langle W, (\geqslant_a)_{a\in\mathcal{A}}, \Delta, I\rangle$ where

- W is the set of states, or worlds;
- $(\geqslant_a)_{a\in\mathcal{A}} \subseteq W \times W$ are the plausibility relations on W, one for each agent, that are converse well-founded, locally connected preorders;
- Δ is the domain of interpretation;
- and I is an *interpretation function* such that $I(w) = \cdot^{I_w}$ and $\cdot^{I_w} : \mathcal{C} \rightarrow \mathcal{P}(\Delta)$, where it holds that $\top^{I_w} = \Delta$.

The semantics of DEOL is equivalent to that of DEL except that we now have instance classifications $C(o)$ and class relations $C \sqsubseteq D$ and $C \oplus D$ that are interpreted by I.

Definition 17 (Satisfiability for DEOL) *Satisfiability* for DEOL extends that of DEL (Definition 14), with an interpretation function I instead of the valuation V, by:

$$\begin{aligned}
&\mathcal{M}, w \models C(o) && \text{iff } o^{I_w} \in C^{I_w}\\
&\mathcal{M}, w \models C \sqsubseteq D && \text{iff } C^{I_w} \subseteq D^{I_w}\\
&\mathcal{M}, w \models C \oplus D && \text{iff } C^{I_w} \cap D^{I_w} = \varnothing
\end{aligned}$$

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 $\Gamma \vDash \phi$) if every pointed model $\langle \mathcal{M}, w \rangle$ satisfying all formulas of Γ, also satisfy ϕ.

DEOL can be seen as a general purpose extension of DEL that could be used to model other games. However, it has been designed to apply to ARG, hence it may require further extensions.

5 Translation

In the previous section, we have introduced Dynamic Epistemic Ontology Logic as a formal framework for knowledge and belief evolution in multi-agent systems that make use of ontologies. We now apply this framework to the Alignment Repair Game. More precisely, we define a translation from ARG states to DEOL theories (§5.1) and from adaptation operators for ARG to dynamic upgrades on DEOL (§5.2), see Figure 3. These translations are labeled by τ and δ, respectively.

This enables us to define and prove the correctness, redundancy and completeness of the adaptation operators in the remainder of this paper.

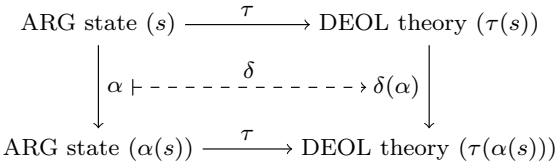

Fig. 3 The translation from ARG states (s) to DEOL theories (τ) and from adaptation operators (α) to dynamic upgrades (δ).

5.1 Translation of ARG states as epistemic axioms

To translate an ARG state to DEOL, we translate the axioms of each agent's ontology as *knowledge* for this agent and each correspondence of alignments between two agents' ontologies as *beliefs* shared by both agents.

Definition 18 (Translation τ) The translation τ from ARG states to DEOL theories is defined by:

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

Example 4 (Running example) In the running example, this means that the axioms $SS_b \sqsubseteq Small_b$, $Small_b \sqsubseteq \top_b$, $TS_b(\blacktriangle)$, etc, of ontology $\mathcal{O}_b$ become $K_b(SS_b \sqsubseteq Small_b)$, $K_b(Small_b \sqsubseteq \top_b)$, $K_b(TS_b(\blacktriangle))$, etc., in DEOL and the correspondences

$\langle Black_a, Small_b, \sqsupseteq\rangle$, $\langle \top_a, \top_b, \sqsupseteq\rangle$ of the alignment A_{ab} and $\langle \top_b, \top_a, \sqsupseteq\rangle$ of A_{ba} become $B_a(Black_a \sqsupseteq Small_b) \wedge B_b(Black_a \sqsupseteq Small_b)$, $B_a(\top_a \sqsupseteq \top_b) \wedge B_b(\top_a \sqsupseteq \top_b)$ and $B_a(\top_a \sqsubseteq \top_b) \wedge B_b(\top_a \sqsubseteq \top_b)$.

Figure 4 depicts the theory $\tau(s)$ from the perspective of agent a. That is, any ϕ such that $K_a\phi$ is drawn in black, and any γ such that $B_a\gamma$ is drawn in blue.

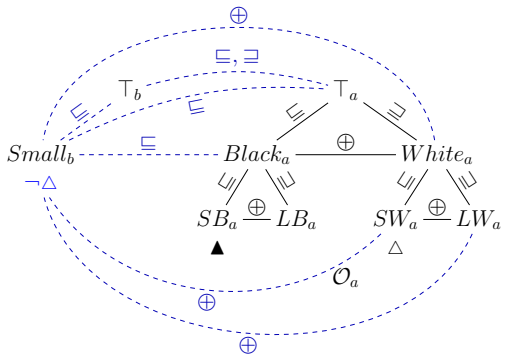

Fig. 4 Initial knowledge (solid black lines) and belief (dashed blue lines) of agent a in the Running example.

5.1.1 Faithfulness

Before we discuss the translation of the adaptation operators, a natural question to ask is: is the translation of an ARG state faithful to this ARG state? More precisely, can anything entailed independently from an ontology or an alignment also be entailed by the transformation as knowledge and beliefs of the agent? If the answer is yes, the translation is consequence preserving.

Proposition 2 (Consequence preservation) *Let s be an ARG state for a set of agents $\mathcal{A}$, then, for each agent $a \in \mathcal{A}$,*

$$\forall\phi: \ \textit{if } \mathcal{O}_a \models \phi \textit{ then } \tau(s) \models K_a\phi \tag{4}$$

and

$$\forall\gamma: \ \textit{if } A_{ab} \models_a \gamma \textit{ then } \tau(s) \models B_a\gamma \tag{5}$$

where the left-hand $\models$ concerns entailment by ontologies and alignments (where $\models_a$ is entailment restricted to agent a) and the right-hand $\models$ concerns entailment in DEOL.

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 2 shows that the translation preserves, modulo modalities, the information that agents have. Given the structure of the translation, it does not seem to introduce arbitrary information. But do the reverse of these statements, let us call them strict adherence, hold?

Let us first consider the reverse of Equation 4. Whenever $\tau(s)$ has a model, i.e. whenever it is consistent, it seems natural that this should hold. However, there are cases where $\tau(s)$ does not have a model – so that anything can be deduced – but the ontologies are consistent – so that not everything can be deduced. This occurs, in particular, when ARG states are locally inconsistent.

Moreover, the converse also holds: whenever $\tau(s)$ has a model, then s must be locally consistent.

Proposition 3 (Local consistence 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.*

Figure 5 shows how we can build a DEOL model $\langle \mathcal{M}^s, w^s \rangle$ for a locally consistent ARG state s with three agents. Essentially, the way to construct $\mathcal{M}^s$ is to create a world that satisfies all the ontologies (this is well-defined because the classes of the ontologies are disjoint) and 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, the two worlds on top for agent a) and believes her alignments (they are true in the most plausible world, the world on top right for agent a).

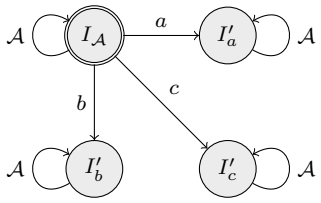

Fig. 5 A model of the ARG state with three agents, $\mathcal{A} = \{a, b, c\}$. We write $I_{\mathcal{A}}$ as abbreviation for the interpretation that assigns to classes $C \in \mathcal{C}_a$ and objects of $\mathcal{D}$ the standard interpretation I_a (Definition 25) for agent a.

Thus, local consistency is preserved from ARG to DEOL and vice versa and there is a procedure to construct a DEOL model of a locally consistent ARG state s such that it satisfies $\tau(s)$. Note that in case of locally consistent but globally inconsistent state, it is possible for the DEOL translation to have a model. Consider Example 5.

Example 5 Let s be the ARG state depicted in Figure 6 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 $C' \sqsupseteq A'$ and $C \sqsupseteq A'$, whereas $C \oplus C'$ and none of these classes can be empty. However,

since the alignments are locally consistent, there exists a model by Proposition 3. 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, agent a only has access to the blue and red alignments, but not green, in Figure 6, agent b to the blue and green ones and agent c to the red and green ones.

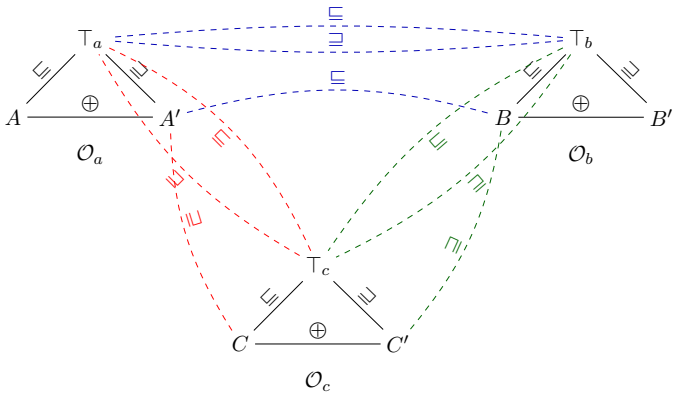

Fig. 6 The ontologies (black) of agent a (left), agent b (right) and agent c (below) and the alignments (blue, red or green, and dashed) between them that is globally inconsistent.

In that sense, DEOL is very faithful to the ARG agents who do not perform any global reasoning.

Whenever $\tau(s)$ has a model, the reverse of Equation 4 holds. We call this strict knowledge adherence: all knowledge of an agent is already entailed by her ontology.

Proposition 4 (Strict knowledge adherence) *Let s be an ARG state locally consistent for a set of agents $\mathcal{A}$, then, for each agent $a \in \mathcal{A}$,*

$$\forall \phi : \textit{if } \tau(s) \models K_a\phi \textit{ then } \mathcal{O}_a \models \phi \tag{6}$$

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

Yet, the reverse for beliefs and alignments, Equation 5, does not hold. This is because, on 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. On ARG, however, agents consider alignments one by one and do not combine the correspondences. Hence some beliefs entailed by the translation may not be entailed by the alignment. Consider Example 6.

Example 6 Let s be the ARG state in Figure 6 with

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

Then $\tau(s) \vDash B_b(B \sqsupseteq A')$ and $\tau(s) \vDash B_b(C' \sqsupseteq B)$, so that $\tau(s) \vDash B_b(C' \sqsupseteq A')$. But it is clear that $A_{ba} \nvDash_b C' \sqsupseteq A'$ nor $A_{cb} \nvDash_b C' \sqsupseteq A'$ because $A', C' \notin \mathcal{C}_b$.

In summary, the translated ARG states are faithful to the initial ARG state with the difference that agents in DEOL reason globally on agent alignments. This is present in both the restriction of using locally consistent states and the absence of strict belief adherence

Now that we have a faithful translation of ARG states to DEOL theories, we consider translating the adaptation operators.

5.2 Translation of adaptation operators as dynamic upgrades

During the gameplay of ARG, agents communicate with each other and through their communication, they learn new information. From a DEL standpoint, 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 (see Definition 10). With a formal model of ARG, we can now ask the question: how do these acts change the knowledge and beliefs of the agents in the DEOL representation of ARG? And are the adaptation operators as defined sufficient and complete to account for these changes?

In order to answer these questions, we translate the communication taking place in ARG, the adaptation operators, to dynamic upgrades on DEOL (according to Figure 3). That is, one round of ARG, as defined in Definition 10 with adaptation operator α applied to correspondence $C_a R C_b$ and object o, is a theory transformation $\delta(\alpha[\langle C_a, C_b, \sqsupseteq\rangle, o])$ adding axioms for the communication that occurs.

Definition 19 (ARG Dynamics in DEOL) Let $T = \tau(s)$ be the DEOL theory that is the translation of an ARG state s. Then $\delta(\alpha[\langle C_a, C_b, \sqsupseteq\rangle, o]) : T \to T'$ is a theory transformation, with adaptation operator α, correspondence $\langle C_a, C_b, \sqsupseteq\rangle$ and object o, where T' is defined as:

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

where $d(\alpha[\langle C_a, C_b, \sqsupseteq\rangle, o])$ is a conservative upgrade of the correspondence deleted or added by the adaptation operator α applied to $\langle C_a, C_b, \sqsupseteq\rangle$ with object o (see Definition 20). We will also write $d(\alpha)$ whenever the correspondence and object are clear from the context.

In the following we will also write $T^{\delta(\alpha)}$ for the transformation $\delta(\alpha)$ applied to T or simply $T^{!C_b(o)}$ ($T^{!C_b(o),d(\alpha)}$, respectively).

On a DEOL model $\mathcal{M}$ of T, the dynamic upgrades defined above (Definition 19) act as model transformations $!C_b(o)$ and $!C_b(o); d(\alpha[\langle C_a, C_b, \sqsupseteq\rangle, o])$, respectively, where the latter is a sequence of upgrades applying first $!C_b(o)$ to $\mathcal{M}$ followed by $d(\alpha[\langle C_a, C_b, \sqsupseteq\rangle, o])$. Thus, $\mathcal{M}$ is transformed to $(\mathcal{M}^{!C_b(o)})^{d(\alpha[\langle C_a, C_b, \sqsupseteq\rangle, o])}$.

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 trustworthyness of the statements. In ARG, first communication occurs and then belief

reorganisation. $C_b(o)$, which is communicated in step 2 of the game (Definition 10) by agent b, is known to agent b because this is part of her ontology. This means that 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. Then, 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. Therefore the corresponding DEOL upgrade for this correspondence is not an announcement but a conservative upgrade – it is soft information, a belief.

Definition 20 (Adaptation Operators as Dynamic Upgrades) Let a correspondence $\langle C_a, C_b, \sqsupseteq\rangle$, an object o, and an adaptation operator α, 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 $M_b(C_b, o) = \{C'_b \in mgcx_b(C_b, o) \mid \nexists\langle C'_a, C'_b, \sqsupseteq\rangle \in A_{ab}\}$.

In the following, we also write $C_a \sqsupset\!\!\!\sqsubset C_b$ for $\neg(C_a \sqsupseteq C_b)$.

Again, as was the case for the adaptation operators on ARG states, 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 seen directly from Definition 20.

Example 7 (Running example - 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 4 can be compared to Figure 7 for an overview of the changes to the epistemic-doxastic state of agent a.

Example 8 (Running example - Failure) If instead ARG is played with △, the round is a failure. Agent b announces $!Small_a(△)$ 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.

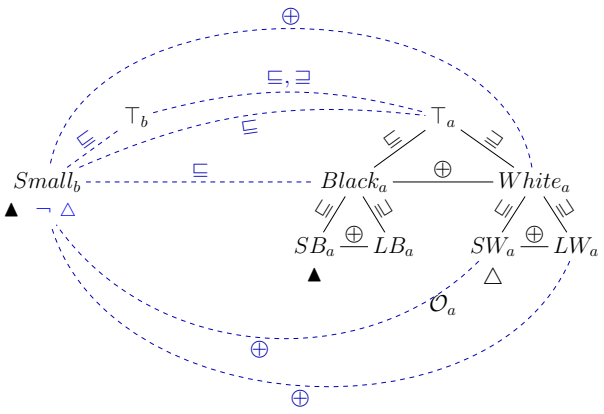

Fig. 7 The knowledge (solid black) and belief (dashed blue) of agent a of Example 2 after the announcement $!Small_b(▲)$.

However, this is not the only revised belief. The contradicted initial beliefs turn into knowledge of their negation. For example, $B_a(\neg Small_b(△))$ becomes $K_a(Small_b(△))$ after the announcement. Compare also Figure 4 and Figure 8 for an overview of the changes to the epistemic-doxastic state of agent a.

According to ARG, an adaptation operator is applied, which results in an updated alignment as explained in Example 3. These correspondences 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. We will see why in the next section.

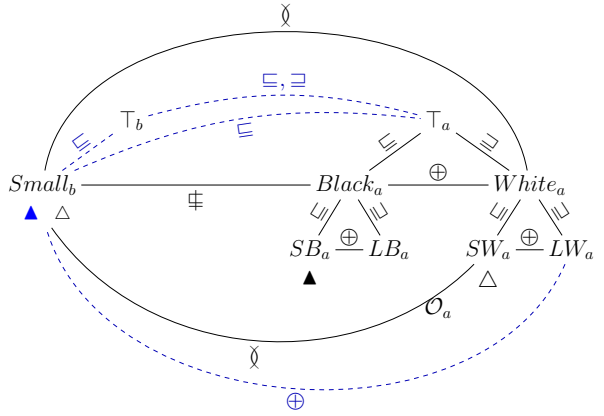

Fig. 8 The knowledge (solid black) and belief (dashed blue) of agent a of Example 2 after the announcement $!Small_b(△)$.

The drawback of modelling 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 a slightly improved version of ARG (with relaxation) brings agents to a fully consistent state even when starting with an inconsistent one [13]. This is not possible in the DEOL translation because it is restricted to consistent states.

6 Formal properties of the Adaptation Operators

With the formal representation of ARG in DEOL we can explore the correctness, redundancy and completeness of the operators. In particular, we show that all the adaptation operators but `add` are correct, `delete`, `addjoin` and `refine` are partially redundant and all adaptation operators are incomplete. To visualize this, we use the diagram in Figure 3.

6.1 Correctness

To show that the adaptation operators are correct, we show that the diagram of Figure 3 commutes. More precisely, that, from an ARG state s, by moving down (α) and then right (τ) we obtain DEOL axioms that are also obtained by moving first right (τ) and then down ($\delta(\alpha)$).

Definition 21 (Correctness) An adaptation operator α is correct if and only if $\forall s$: $(\tau(s))^{\delta(\alpha)} \models \tau(\alpha(s))$.

Proposition 5 (Correctness) *The adaptation operators* `delete`*,* `addjoin`*,* `refine` *and* `refadd` *are correct.*

Yet, the adaptation operator `add` is not correct because it does not take into account whether the most specific superclass of C_a is consistent with the object o. Moreover, if it is consistent, `add` is equivalent to `addjoin`.

Proposition 6 (Incorrectness of `add`) *The adaptation operator* $\alpha =$ `add` *is incorrect, i.e.* $\exists s : (\tau(s))^{\delta(\texttt{add})} \not\models \tau(\texttt{add}(s))$*, and* $\forall s$ *such that* $(\tau(s))^{\delta(\texttt{add})} \models \tau(\texttt{add}(s))$*:* $\texttt{add}(s) = \texttt{addjoin}(s)$.

Proposition 6 is in line with initial predictions and experimental results [13]: `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 these results, it is clear that for a logical agent, `add` should be abandoned.

6.2 Redundancy

The redundancy of some operators in the running example is not a coincidence. For logical agents, i.e. DEOL agents, some adaptation operators are redundant for *every* ARG state: `delete` and `addjoin` are redundant with respect to agent a and `refine` is redundant with respect to agent b. Before we define this redundancy with respect to one agent (*partial redundancy*), let us first consider what it means for an operator to be redundant (with respect to *both* agents). An adaptation operator α is redundant if and only if solely applying $!C_b(o)$ to the DEOL translation of s is already sufficient to obtain an interpretation of the DEOL translation of $\alpha(s)$.

Definition 22 (Redundancy) An adaptation operator α is *redundant* if and only if $\forall s$: $(\tau(s))^{!C_b(o)} \models \tau(\alpha(s))$.

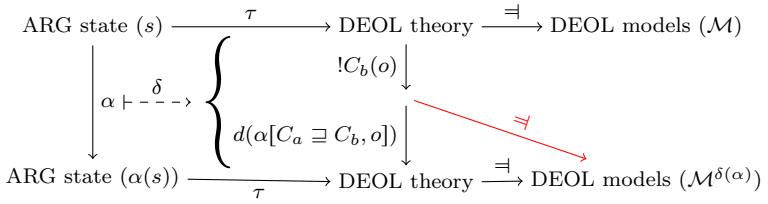

Fig. 9 The translation from ARG states (s) to DEOL theories (τ), interpreted on DEOL models ($\mathcal{M}$), and from adaptation operators (α) to dynamic upgrades (δ) for redundant adaptation operators. Hence, $\mathcal{M}^{\delta(\alpha)}$ is already a model of $\tau(s) \cup !C_b(o)$.

The adaptation operators discussed here are not redundant, but *partially redundant*. This means that they are redundant with respect to one agent. To prove redundancy, we show that the knowledge and belief of this agent are invariant to the application of the adaptation operator. In fact, because adaptation operators only alter the beliefs of agents, it suffices to show partial redundancy by showing that the beliefs of that agent remain unchanged.

Definition 23 (Partial Redundancy) An adaptation operator α is *partially redundant* for agent a if and only if $(\tau(s))^{!C_b(o)} \models B_a\phi$ implies $\tau(\alpha(s)) \models B_a\phi$ for each ARG state s and each ϕ in $\mathcal{L}_{DEOL}$.

Proposition 7 (Partial redundancy) *The adaptation operators* `delete` *and* `addjoin` *are partially redundant with respect to agent a, and* `refine` *is partially redundant with respect to agent b.*

However, none of the adaptation operators is redundant with respect to both agents. 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.

6.3 Incompleteness

Finally, we consider completeness of the adaptation operators: do the operators capture all the information that can be learned? Intuitively, this is proven by

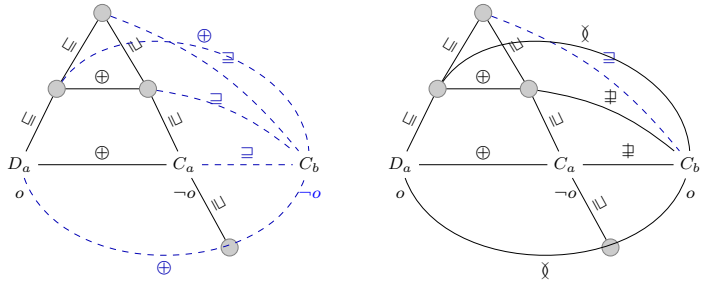

Fig. 10 The knowledge (solid black) and beliefs (dashed blue) of agent a before (left) and after (right) the announcement $!C_b(o)$.

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.

To show that the adaptation operators are complete, we show that the diagram of Figure 3 commutes in the reverse direction compared to correctness. More precisely, that, from an ARG state s, by moving right (τ) and then down ($\delta(\alpha)$) we obtain DEOL axioms that are also obtained by moving first down (α) and then right (τ).

Definition 24 (Completeness) An adaptation operator α is complete if and only if $\forall s$: $\tau(\alpha(s)) \models (\tau(s))^{\delta(\alpha)}$.

Proposition 8 (Incompleteness) *All adaptation operators are incomplete.*

6.4 Discussion

Experiments have shown that ARG works quite well in practice: through application of the adaptation operators, agents can reach successful communication [11, 13]. Furthermore, the adaptation operators are ARG state preserving (Property 1), produce safe and entailed alignments [13] and work even with globally inconsistent networks of ontologies (Figure 6). Why is that despite the lack of formal properties? In short, this is because agents do not need to be logical to be successful.

For example, the incompleteness of the adaptation operators is based on the fact that ARG agents do not memorize the failure of the correspondence with the drawn object. They forget to which class the other agent classifies the drawn object – yet, this is deduced knowledge in DEOL through the announcement $!C_b(o)$. Nevertheless, even without this knowledge, agents can communicate successfully. That is, they do not need the concrete examples (the object o 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 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. Hence, one may think that the announcement could have

been avoided in the translation of operators in order to prevent incompleteness and redundancy. However, the main objection is that the current translation is indeed faithful to the ARG game: agents announce to which class the object belongs, they use this fact to improve their alignment and then forget it.

Another alternative would be to concentrate on the alignment only instead of measuring the formal properties with respect to the full knowledge and belief of the agent. The operators may then be complete. This remains an open question.

7 Conclusion

In order to understand the formal properties of adaptation operators used in the alignment repair game, we modelled the game into a dynamic epistemic logic, DEOL, with embedded ontological statements. We assessed the faithfulness of this translation 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.

This work bridges a very practical implementation of adaptive agents used in simulations to a dynamic epistemic model of logical agents. In spite of, or because of, the simplicity of ARG, this revealed more challenging than expected. Developing a tighter connection requires addressing fundamental issues that we discuss hereafter.

There are two ways to interpret our results: either the ARG agents use sub-logical behavior, or our logical modeling is too strong and not faithful to them, although the results in Section 5.1.1 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 ARG agents that reason more faithfully to the DEOL logic. Although this was not the goal of the initial experiments, providing agents with more logical reasoning power can be considered. The opposite direction consists in bringing the logic closer to the agents. In particular, the proof of incompleteness draws on the inability of agents to remember individual cases because they focus on general knowledge (they draw general conclusions but forget the instances that led them to these conclusions). Introducing forgetting operations may help reaching a closer model [32, 33].

There are, however, other differences between ARG agents and their logical modeling in DEOL that would be worth investigating. A first one is the difference between local and global reasoning. So far we have compared the encoding of ontologies and alignments locally, i.e. one by one. This is a correct way to model agents because they indeed can only reason locally. However, as shown in Section 5.1.1, this is another place where it could be shown that the logical model is more powerful (can deduce more) than agents. Another difference comes from the logical model to be defined in function of a unique signature. In open environments, in which agents should be able to evolve, it is natural that agents are not aware of the signatures of other agents and discover them as their knowledge evolves. Yet, if this is not necessary, they will not be aware of their whole signatures. The current

logical model assumes a fixed set of symbols shared by all agents. It would be a desirable feature for a model to drop this assumption and to model unawareness of agents and the way agents can raise their awareness progressively [15,32,33,30, 6].

This work may also be linked to belief revision. Indeed, the announcements cause agents to change their beliefs so that consistency is preserved. However, they do *not* do it in a minimal way. The ARG adaptation operators, besides `delete`, have been designed to address this problem [13]: the added correspondences were entailed by the deleted one. Although ARG has not been studied under the light of belief revision, this has been the case of dynamic logic [4,27]. Hence, this model may be used to find out if the adaptation operators are correct belief revision operations.

Beyond the specific case of the ARG game, this work provides a methodology to assess other games within dynamic epistemic logic. Indeed, it suggests how to translate such games (ontologies and gameplay) to DEOL and it defines properties that could be satisfied (correctness, completeness and redundancy). Of course, as discussed before, these are properties of both the game and the translation, hence one must be cautious about what the translation preserves (faithfulness).

A final perspective of this work is that Dynamic Epistemic Logic seems a promising framework for modelling cultural evolution or, at least, cultural knowledge evolution. Indeed, it offers ways to express agent beliefs and to announce, i.e. communicate, assertions. It can also constrain consistency and finally belief revision methods have been tied to it. This may be the basis to logic-based adaptation operators (replacing the DEOL upgrades corresponding to the adaptation operators used in this work).

Acknowledgements This work has been partially supported by MIAI @ Grenoble Alpes (ANR-19-P3IA-0003).

A Proofs

A.1 Properties of ARG ontologies

Lemma 2 *For any ARG ontology $\mathcal{O}$ of signature $\langle \mathcal{C}, \mathcal{D} \rangle$, $\forall C \in \mathcal{C}$, there exists a unique finite chain $C_0, \dots C_n$; $\forall i \in [1, n]$, $C_i \sqsubseteq C_{i+1} \in \mathcal{O}$, $C_0 = C$ and $C_n = \top$.*

Proof This is a consequence Definition 2: For any $C \in \mathcal{C}$, either $C = \top$, and then the property is satisfied with the chain reduced to $\top$ ($n = 0$). It is unique because there exist no $\top \sqsubseteq D \in \mathcal{O}$ (Definition 2(1)). Otherwise, there exists a unique C_1 such that $C \sqsubseteq C_1 \in \mathcal{O}$ (Definition 2(2)). This reasoning can be iterated until finding $C_n = \top$. This is warranted because $\mathcal{O}$ is finite (Definition 2) and the graph of $\sqsubseteq$ in $\mathcal{O}$ contains no cycle (Definition 2(5)). □

Definition 25 (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:

$$\text{if } C \sqsubseteq D \in \mathcal{O} \text{ and } o \in \mathcal{D} \text{ such that } o \in C^{\hat{I}_k} \text{ and } o \notin D^{\hat{I}_k} \text{ then } o \in D^{\hat{I}_{k+1}} \tag{7}$$

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 also say *standard interpretation* for a class C for the interpretation of C in the standard model.

Proof (Proposition 1) Let $\mathcal{O}$ be an ARG ontology over a signature $\langle \mathcal{C}, \mathcal{D} \rangle$ and let $\hat{I}$ be the standard model of $\mathcal{O}$.

The standard model is well defined because both $\mathcal{O}$ and $\mathcal{D}$ are finite and therefore (7) can only be applied a finite number of times until $n \in \mathbb{N}$ is reached such that $\hat{I}_n = \hat{I}_m$ for every $m \geqslant n$. This means that the sequence $\{\hat{I}_k\}_{k \geqslant 0}$ converges to $\hat{I}_n$, which is defined as $\hat{I}$.

Let us prove that $\hat{I}$ is a model of $\mathcal{O}$. Let $C \sqsubseteq D \in \mathcal{O}$. Assume that $\hat{I} \not\models C \sqsubseteq D$. Then, there exists $o \in \mathcal{D}$ such that $o \in C^{\hat{I}}$ and $o \notin D^{\hat{I}}$. This means that (7) can be applied to extend $\hat{I} = \hat{I}_n$ to $\hat{I}_{n+1}$ by adding o to $\hat{I}_{n+1}$. But this contradicts the fact that $\hat{I}$ is defined $\hat{I}_n$ such that for every $m \geqslant n$: $\hat{I}_n = \hat{I}_m$. So in particular, $\hat{I}_{n+1} = \hat{I}_n$. Therefore, $\hat{I} \models C \sqsubseteq D$.

Let $C \oplus D \in \mathcal{O}$ and assume, towards contradiction, that $\hat{I} \not\models C \oplus D$. This means that there exists $o \in \mathcal{D}$ such that $o \in C^{\hat{I}}$ and $o \in D^{\hat{I}}$. By construction of $\hat{I}$, there are four possibilities:

1. $o \in C^{\hat{I}_0}$ and $o \in D^{\hat{I}_0}$. In this case, $C(o), D(o) \in \mathcal{O}$. Therefore, $\mathcal{O} \not\models C \oplus D$, which contradicts the fact that $C \oplus D \in \mathcal{O}$.
2. $o \in C^{\hat{I}_0}$ but $o \notin D^{\hat{I}_0}$. Then, by definition of $\hat{I}$, there exist $D_0, \dots, D_k \in \mathcal{C}$ ($k > 0$) such that $D_k = D$, for $0 \leqslant i \leqslant k-1$: $D_i \sqsubseteq D_{i+1} \in \mathcal{O}$ and $o \in D_0^{\hat{I}_0}$. Then, by (7), for $0 \leqslant i \leqslant k-1$: $o \in D_i^{\hat{I}_i}$, $o \notin D_{i+1}^{\hat{I}_i}$ and $o \in D_{i+1}^{\hat{I}_{i+1}}$. Since $D_0(o) \in \mathcal{O}$ and $D_i \sqsubseteq D_{i+1}$ ($0 \leqslant i \leqslant k-1$) and $D_k = D$, then $\mathcal{O} \models D(o)$. But then, since $C(o) \in \mathcal{O}$, it holds that $\mathcal{O} \not\models C \oplus D$, which contradicts the fact that $C \oplus D \in \mathcal{O}$.
3. $o \in D^{\hat{I}_0}$ but $o \notin C^{\hat{I}_0}$. Then, similarly to the argument above, there exists $C_0, \dots, C_k \in \mathcal{C}$ ($k > 0$) such that $C_k = C$, for $0 \leqslant i \leqslant k-1$: $C_i \sqsubseteq C_{i+1} \in \mathcal{O}$ and $o \in C_0^{\hat{I}_0}$. And therefore, analogous to the case above, we have $D(o) \in \mathcal{O}$ but $\mathcal{O} \models C(o)$ and hence $\mathcal{O} \not\models C \oplus D$, contradicting $C \oplus D \in \mathcal{O}$.
4. $o \notin C^{\hat{I}_0}$ and $o \notin D^{\hat{I}_0}$. Then, again, there exists classes $C_0, \dots, C_k \in \mathcal{C}$ and $D_0, \dots, D'_k \in \mathcal{C}$ ($k, k' > 0$) to infer that $\mathcal{O} \models C(o)$ and $\mathcal{O} \models D(o)$ and hence $\mathcal{O} \not\models C \oplus D$, contradicting $C \oplus D \in \mathcal{O}$.

And therefore it must hold that $\hat{I} \models C \oplus D$.

Finally, let $C(o) \in \mathcal{O}$. Then, by definition of the standard model, $o \in C^{\hat{I}_0}$. But then, because for each $k \in \mathbb{N}$, $\hat{I}_{k+1}$ extends $\hat{I}_k$, it must be that $o \in C^{\hat{I}_k}$. Thus $o \in C^{\hat{I}}$ and hence $\hat{I} \models C(o)$. □

We first establish four properties of ARG ontologies which are used as arguments in various proofs:

Lemma 3 *For any ARG ontology $\mathcal{O}$ of signature $\langle \mathcal{C}, \mathcal{D} \rangle$:*
(a) $\forall C \in \mathcal{C}$, $\mathcal{O} \models C \sqsubseteq \top$.
(b) For any two different class names $D, D' \in \mathcal{C} \backslash \{\top\}$, either $\mathcal{O} \models D \sqsubseteq D'$ or $\mathcal{O} \models D' \sqsubseteq D$ or $\mathcal{O} \models D' \oplus D$.
(c) For each $C \in \mathcal{C}$, $\exists o \in \mathcal{D}$; $\mathcal{O} \models C(o)$.
(d) For all classes $C, C' \in \mathcal{C}$, if $\mathcal{O} \models C \sqsubseteq C'$ and $\mathcal{O} \models C' \sqsubseteq C$, then $C = C'$.

Proof (a) $\forall C \in \mathcal{C}$, $\mathcal{O} \models C \sqsubseteq \top$. For all model I of $\mathcal{O}$, $\top^I = \Delta$ and $C^I \subseteq \Delta$, hence $C^I \subseteq \top^I$.
(b) For any two different class names $D, D' \in \mathcal{C} \backslash \{\top\}$, either $\mathcal{O} \models D \sqsubseteq D'$ or $\mathcal{O} \models D' \sqsubseteq D$ or $\mathcal{O} \models D' \oplus D$. We know that (Definition 2(5)) $\exists C_0, \ldots C_m$; $\forall i \in [1, m]$, $C_i \sqsubseteq C_{i-1} \in \mathcal{O}$, $C_0 = \top$ and $C_m = D$ and $\exists C'_0, \ldots C'_n$; $\forall j \in [1, n]$, $C'_i \sqsubseteq C'_{i-1} \in \mathcal{O}$, $C'_0 = \top$ and $C'_n = D'$. There is a rank $k \leqslant \min(m, n)$, such that $\forall i \in [0, k]$, $C_i = C'_i$ (at the minimum this is true for $k = 0$). If $k = m$, then $C_k = D$ and $\mathcal{O} \models D \sqsubseteq D'$; if $k = n$, then $C'_k = D'$ and $\mathcal{O} \models D' \sqsubseteq D$; otherwise $C_{k+1} \neq C'_{k+1}$. This means that (Definition 2(3)) $C_k \oplus C'_k \in \mathcal{O}$. But it can easily be checked that for each model I of $\mathcal{O}$, $D^I \subseteq C^I_{k+1}$, $D'^I \subseteq C'^I_{k+1}$, and $C^I_{k+1} \cap C'^I_{k+1} = \varnothing$. Hence, $D^I \cap D'^I = \varnothing$, so $\mathcal{O} \models D \oplus D'$.
(c) For each $C \in \mathcal{C}$, $\exists o \in \mathcal{D}$; $\mathcal{O} \models C(o)$. Either (Definition 2(3)) $\exists o \in \mathcal{D}$; $C(o) \in \mathcal{O}$, thus $\mathcal{O} \models C(o)$, or $\exists C' \in \mathcal{C}$; $C \sqsubseteq C' \in \mathcal{O}$. In the latter case, since $\mathcal{C}$ is finite and (4) the graph made by $\sqsubseteq$-axioms in $\mathcal{O}$ is non circular, $\exists C_0, \ldots C_m$; $\forall i \in [1, m]$, $C_i \sqsubseteq C_{i-1} \in \mathcal{O}$, $C_0 = C$ and (Definition 2(3)) $C_m(o) \in \mathcal{O}$. This entails that, for each model I of $\mathcal{O}$, $o^I \in C^I_m \subseteq \cdots \subseteq C^I$, hence $\mathcal{O} \models C(o)$.
(d) For all classes $C, C' \in \mathcal{C}$, if $\mathcal{O} \models C \sqsubseteq C'$ and $\mathcal{O} \models C' \sqsubseteq C$, then $C = C'$. This can be rephrased as: $\forall C, C' \in \mathcal{C}$, if $C \neq C'$, then it is not possible to have $\mathcal{O} \models C \sqsubseteq C'$ and $\mathcal{O} \models C' \sqsubseteq C$. We know that (Definition 2(5)) $\exists C_0, \ldots C_m$; $\forall i \in [1, m]$, $C_i \sqsubseteq C_{i-1} \in \mathcal{O}$, $C_0 = \top$ and $C_m = D$ and $\exists C'_0, \ldots C'_n$; $\forall j \in [1, n]$, $C'_i \sqsubseteq C'_{i-1} \in \mathcal{O}$, $C'_0 = \top$ and $C'_n = D'$. There is a rank $k \leqslant \min(m, n)$, such that $\forall i \in [0, k]$, $C_i = C'_i$ (at the minimum this is true for $k = 0$). Since $C \neq C'$, then $k \neq m$ or $k \neq n$. Hence (Definition 2(3)) $\exists! \langle D, D' \rangle \in \mathcal{C} \times \mathcal{C}$; $D \sqsubseteq C_k \in \mathcal{O}$, $D' \sqsubseteq C_k \in \mathcal{O}$, and $D \oplus D' \in \mathcal{O}$. Consider that $k \neq m$ and $k \neq n$, then for all models I of $\mathcal{O}$, $C^I \subseteq D^I$ and $C'^I \subseteq D'^I$, hence $C^I \cap C'^I = \varnothing$, but (Lemma 3(c)) $C^I \neq \varnothing$, thus $\mathcal{O} \not\models C \sqsubseteq C'$. Now, consider that $k = n$ and $k \neq m$, then (Definition 2(3)) $\exists! \langle D, D' \rangle \in \mathcal{C} \times \mathcal{C}$; $D \sqsubseteq C_k \in \mathcal{O}$, $D' \sqsubseteq C_k \in \mathcal{O}$, and $D \oplus D' \in \mathcal{O}$. Moreover, (Lemma 3(c)) $\exists o, o' \in \mathcal{D}$; $\mathcal{O} \models C(o)$, $\mathcal{O} \models D'(o')$. This means that for all models I of $\mathcal{O}$, $o^I \in C^I \subseteq D^I$, $o'^I \in D'^I \subseteq C^I_k = C'^I$ and $C^I \cap D'^I = \varnothing$, hence $o'^I \notin C^I$, thus $\mathcal{O} \not\models C' \sqsubseteq C$. □

Clause (d) shows that $\sqsubseteq$ is a partial order on $\mathcal{C}$ for $\mathcal{O}$ (it inherits transitivity and reflexivity from sets). The lemma is stated for consistent ARG ontologies but by their constraints, any ARG ontology is consistent.

Proof (Lemma 1)
(a) Let $Sup(C) = \{D \in \mathcal{C} \backslash \{C\}; \mathcal{O} \models C \sqsubseteq D\}$, the set of C subsumer names. This set is non empty because (Lemma 3(a)) it contains $\top$. That $Sup(C)$ has no most specific element would mean that it contains at least two different class names D and D' such that neither $\mathcal{O} \models D \sqsubseteq D'$, nor $\mathcal{O} \models D' \sqsubseteq D$. Hence, (Lemma 3(b)) $\mathcal{O} \models D \oplus D'$. However, (Lemma 3(c)) there exists $o \in \mathcal{D}$ such that $\mathcal{O} \models C(o)$. That would mean that for all models I of $\mathcal{O}$, $o^I \in C^I \subseteq D^I$, $o^I \in C^I \subseteq D'^I$ and $D^I \cap D'^I = \varnothing$. There cannot be such models.
(b) Let $Cl(o) = \{D \in \mathcal{C}; \mathcal{O} \models D(o)\}$ the set of classes to which o belongs. This set is non empty as it contains $\top$, indeed, for all models I of $\mathcal{O}$, $o^I \in \top^I$. That $Cl(o)$ has no most specific element would mean that it contains at least two different class names D and D' such that neither $\mathcal{O} \models D \sqsubseteq D'$, nor $\mathcal{O} \models D' \sqsubseteq D$. Hence, (Lemma 3(b)) $\mathcal{O} \models D \oplus D'$. That would mean that for all models I of $\mathcal{O}$, $o^I \in D^I \cap D'^I$ and $D^I \cap D'^I = \varnothing$. There cannot be such models.
(c) Consider $Sup(C, o) = (Sup(C) \cup \{C\}) \cap Cl(o)$. This set contains at least $\top$. Since $Sup(C, o)$ is a set of classes to which o belong, the same argument as for the previous property can be used. If it contains C (i.e. $\mathcal{O} \models C(o)$), then $C = mscc(o, C)$.
(d) Let $Sub(C) = \{C' \in \mathcal{C} | \mathcal{O} \models C' \sqsubseteq C\}$, $mgcx(C, o)$ is defined as the maximal elements ($\sqsubseteq$-wise) of the elements of $Sub(C)$ such that $\mathcal{O} \not\models C'(o)$. This is always well-defined as these constraints are well-defined. If $Sub(C)$ is empty (C has no subclasses) or none of its elements satisfy the constraint, then $mgcx(C, o) = \varnothing$. Otherwise, $mgcx(C, o)$ is not empty because among those elements of $Sub(C)$ satisfying the constraint, there are maximal ones because $\sqsubseteq$ is a (finite) partial order on $\mathcal{C}$ for $\mathcal{O}$. □

A.2 Proof of adaptation operators

Proof (Property 1) Consider without loss of generality that $c = \langle C_a, C_b, \sqsupseteq \rangle \in A_{ab}$. Only alignment A_{ab} of the initial ARG state is modified, hence it is necessary to verify that the modifications are well-defined and that their application to A_{ab} returns an ARG alignment. The only deletion concerns the faulty correspondence $\langle C_a, C_b, \sqsupseteq \rangle$. Since it belongs to A_{ab}, its deletion is always possible. The only way it could prevent the modified alignment to be an ARG alignment, would be if $C_a = \top_a$ and $C_b = \top_b$ (because A_{ab} is an ARG alignment, there can be only one alignment arriving at class $\top_b$). But this would mean that $\mathcal{O}_a \not\models \top_a(o)$ which is not possible because in any model $\langle \cdot^I, \Delta \rangle$ of $\mathcal{O}_a$, $o^I \in \Delta$ and $\top_a^I = \Delta$. Hence, the deletion is well-defined and preserves the ARG state. Hence, `delete` is already well-defined. Concerning the addition, this concerns $\langle msc_a(C_a), C_b, \sqsupseteq \rangle$ (`add`), $\langle mscc_a(o, C_a), C_b \sqsupseteq \rangle$ (`addjoin`), and $\{\langle C_a, C'_b, \sqsupseteq \rangle\}$ (`refine`). First, from Lemma 1, $msc_a(C_a)$ and $mscc_a(o, C_a)$ always exists and are unique which make the operators well-defined. Moreover, because A_{ab} is an ARG alignment, it had only one correspondence $\langle C_a, C_b, \sqsupseteq \rangle$, that has been deleted, then by adding the new correspondence, the new alignment is also an ARG alignment. Next, $mgcx_b(C_b, o)$ is well-defined and a correspondence $\langle C_a, C'_b, \sqsupseteq \rangle$ is added by `refine` only if there is no other conflicting correspondence $\langle C'_a, C'_b, \sqsupseteq \rangle$ in A_{ab}. Hence, the result of `refine` is an ARG alignment. Finally, since `refadd` adds correspondences from different classes of $\mathcal{O}_b$ (C_b for `addjoin` and the elements of $mgcx_b(C_b, o)$ for `refine` and $C_b \notin mgcx_b(C_b, o)$ otherwise the failure would not have occured), there cannot be two correspondences from the same class, hence this generates an ARG alignment. □

A.3 Proofs of faithfulness

Proof (Proposition 2)

Proof of Equation 4

We consider two cases: (1) $\tau(s)$ has a model and (2) $\tau(s)$ has no model. Whenever (2), Equation 4 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 18 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 $\tau(s) \models K_a\phi$.

Proof of Equation 5

Again, we consider the two cases: (1) $\tau(s)$ has a model and (2) $\tau(s)$ has no model, and whenever (2), Equation 5 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 18), 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 18). 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 $\tau(s) \models B_a\gamma$. □

Definition 26 (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:

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

And the interpretation I^s assigns:

1. to each world v and each object name $o \in \mathcal{D}$ the object o: i.e. $I^s_v(o) = o$,
2. to w^s and class name $C \in \mathcal{C}_a$ the *standard interpretation* (Definition 25) $\hat{I}_a$ of $\mathcal{O}_a$: i.e. $C^{I^s_{w^s}} = C^{\hat{I}_a}$,
3. and to w_a an interpretation I'_a such that $I_a \subseteq I'_a$ for $I_a \models_a \bigcup_{b \in \mathcal{A}\backslash\{a\}} A_{ab}$, i.e. $C^{I^s_{w_a}} = C^{I'_a}$.

The DEOL model in Definition 26 is constructed in such a way that for each agent a, there is a most plausible world w_a reached from w^s such that the interpretation at w_a is exactly the extension I'_a of the interpretation I_a that locally satisfies all the correspondences of all the alignments involving a. I'_a is guaranteed to exists because s is locally consistent. However, I'_a is not unique. There can be multiple interpretations I'_a that satisfy the requirement that $I'_a \supseteq I_a$ for some I_a such that $I_a \models_a \bigcup_{b \in \mathcal{A}\backslash\{a\}} A_{ab}$, and hence also multiple DEOL models $\langle \mathcal{M}^s, w^s \rangle$. This is not important, because we just need one. And the existence of one is guaranteed by the fact that s is locally consistent.

Lemma 4 *For all ARG states s, $\langle \mathcal{M}^s, w^s \rangle$ as defined in Definition 26 is a model of $\tau(s)$.*

Proof 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 beliefs her alignments and so $\langle \mathcal{M}^s, w^s \rangle$ is a model of $\tau(s)$. □

Proof (Proposition 3)
(if) Assume that s is an ARG state for a set of agents $\mathcal{A}$ with locally consistent alignments. That 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 $\bigcup_{b \in \mathcal{A}\backslash\{b\}} A_{ab}$. Thus there is an extension I'_a of I_a that locally satisfies all $\gamma \in \bigcup_{b \in \mathcal{A}\backslash\{b\}} A_{ab}$ 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 26, then, by Lemma 4, $\langle \mathcal{M}^s, w^s \rangle$ 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. □

For the proof of Proposition 4, we prove a property of standard interpretations (Definition 25): for two different classes C, D, if the canonical 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 5 *For each locally consistent ARG state s and for each canonical 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 26, 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(4)), 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(3)). 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 (Lemma 2), 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(3)). 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(3)). 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 (Proof of Proposition 4) 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(4)) 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) $\mathcal{O}_a \models C \oplus D$ (Lemma 3(b)). 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 (Lemma 3(b)) $\mathcal{O}_a \models \psi$ 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 (Lemma 3(c)) 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 just as in the previous case no interpretation can accommodate $\psi = C \oplus D$. In order 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 2), also $\mathcal{O}_a \not\models D \sqsubseteq C$ and thus, by Lemma 3(b), $\mathcal{O}_a \models C \sqsubseteq D$. Since s is locally consistent, we can a DEOL model $\langle \mathcal{M}^s, w^s \rangle$ of s (Definition 26) such that $\mathcal{M}^s, w^s \models \tau(s)$ and therefore $\mathcal{M}^s, w^s \models K_a(C \sqsubseteq D)$, i.e. $C^{I_w} \subseteq D^{I_w}$ for each $w \in \{w^s, w_a\}$ and in particular this must hold at 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 5, $D^{I_a} \nsubseteq C^{I_a}$ and thus $D^{I_{w^s}} \nsubseteq C^{I_{w^s}}$. 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. □

A.4 Proofs of correctness, redundancy and incompleteness

Proof (Proposition 5) We do the proof for agent a and adaptation operator `addjoin`. The proof for `delete` now follows because it is entailed by `addjoin`, and the proof for `refine` is symmetric.

Because `addjoin` only adds beliefs, it suffices to show that these beliefs are entailed (we note C_a^o for $mscc_a(o, C_a)$): $(\tau(s))^{!C_b(o);\uparrow(C_a \sqsupseteq C_b \wedge C_a^o \sqsupseteq C_b)} \models B_i(C_a \sqsupseteq C_b) \wedge B_i(C_a^o \sqsupseteq C_b)$. This holds because initially the correspondence is believed, i.e. $\tau(s) \models B_i(C_a \sqsupseteq C_b)$, and the upgrade $!C_b(o); \uparrow (C_a \sqsupseteq C_b \wedge C_a^o \sqsupseteq C_b)$ deletes all the worlds from $\tau(s)$ in which $C_b(o)$ is false and then rearranges the remaining worlds such that the '$C_a \sqsupseteq C_b \wedge C_a^o \sqsupseteq C_b$'-worlds become more plausible than the '$\neg(C_a \sqsupseteq C_b \wedge C_a^o \sqsupseteq C_b)$'-worlds. Because there remain '$C_a \sqsupseteq C_b \wedge C_a^o \sqsupseteq C_b$'-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 $C_a \sqsupseteq C_b$ or $C_a^o \sqsupseteq C_b$ are unchanged. Therefore the beliefs $B_i(C_a \sqsupseteq C_b)$ and $B_i(C_a^o \sqsupseteq C_b)$ are enforced for agents $i \in \{a, b\}$. Hence `addjoin` is correct. □

Proof (Proposition 6) 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 $\delta(\texttt{add}) = !C_b(o); \uparrow (C_a \sqsupseteq C_b \wedge msc_a(C_a) \sqsupseteq C_b)$, 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 $(\tau(s))^{\delta(\texttt{add})} \models K_a(C_b(o)) \wedge K_a(C' \sqsupseteq C_b)$. 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`. □

Proof (Proposition 7) We do the proof for agent a and the adaptation operator `addjoin`. The proof for `delete` now follows because it is entailed by `addjoin`, and the proof for `refine` is symmetric.

Thus we need to show that if $(\tau(s))^{!C_b(o)} \models B_a\phi$ holds, then it is also the case that $\tau(\texttt{addjoin}[\langle C_a, C_b, \sqsupseteq\rangle, o](s)) \models B_a\phi$ holds. Consider a sentence ϕ that is *not* believed by agent a in the state transformed by the adaptation operator, $\tau(\texttt{addjoin}[\langle C_a, C_b, \sqsupseteq\rangle, o](s))$, but is in the original state $\tau(s)$. By construction of the dynamics of the operator `addjoin`, this can only be (1) a belief that is inconsistent with $C_b(o)$ (because the announcement $!C_b(o)$ deletes these worlds), or (2) $C_a \not\sqsupseteq C_b$ (because it is enforced by the conservative upgrade part of the dynamics). But these are also not believed by agent a in $(\tau(s))^{!C_b(o)}$: (1) because $!C_b(o)$ has deleted all these beliefs, and (2) because $\tau(s) \models K_a(\neg(C_a(o)))$ and this knowledge is invariant under the announcement $!C_b(o)$, causing the belief in $C_a \not\sqsupseteq C_b$ to be dropped. Hence, by contraposition, `addjoin` is partially redundant with respect to agent a. In Figure 10 the knowledge and belief of agent a is illustrated before and after the announcement $!C_b(o)$ for an intuition. □

Proof (Proposition 8) Again, consider the knowledge and belief of agent a before and after the announcement $!C_b(o)$, see also Figure 10. 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 $\tau(\alpha(s)) \not\models (\tau(s))^{\delta}$. □

References

1. Aberer, K., Cudré-Mauroux, P., Hauswirth, M.: Start making sense: The chatty web approach for global semantic agreements. J. Web Semant. **1**(1), 89–114 (2003). DOI 10.1016/j.websem.2003.09.001
2. Atencia, M., Schorlemmer, M.: An interaction-based approach to semantic alignment. Journal of Web Semantics **12**, 131–147 (2012). DOI 10.1016/j.websem.2011.12.001
3. 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)
4. Baltag, A., Smets, S.: Dynamic belief revision over multi-agent plausibility models. In: Proceedings of LOFT, vol. 6, pp. 11–24. University of Liverpool (2006)
5. 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)
6. Van den Berg, L., Atencia, M., Euzenat, J.: Unawareness in multi-agent systems with partial valuations. In: Proceedings of the 10th AAMAS workshop on Logical Aspects of Multi-Agent Systems (LAMAS) (2020)
7. 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`
8. Chocron, P., Schorlemmer, M.: Attuning ontology alignments to semantically heterogeneous multi-agent interactions. In: ECAI 2016 - 22nd European Conference on Artificial Intelligence, 29 August-2 September 2016, The Hague, The Netherlands - Including Prestigious Applications of Artificial Intelligence (PAIS 2016), pp. 871–879 (2016). DOI 10.3233/978-1-61499-672-9-871
9. Chocron, P., Schorlemmer, M.: Vocabulary alignment in openly specified interactions. In: Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems, AAMAS 2017, São Paulo, Brazil, May 8-12, 2017, pp. 1064–1072 (2017). URL `http://dl.acm.org/citation.cfm?id=3091275`
10. Chocron, P., Schorlemmer, M.: Vocabulary alignment in openly specified interactions. Journal of Artificial Intelligence Research **68**, 69–107 (2020)
11. Euzenat, J.: First experiments in cultural alignment repair (extended version). In: The Semantic Web: ESWC 2014 Satellite Events - ESWC 2014 Satellite Events, Anissaras, Crete, Greece, May 25-29, 2014, Revised Selected Papers, pp. 115–130 (2014). DOI 10.1007/978-3-319-11955-7_10
12. Euzenat, J.: Revision in networks of ontologies. Artificial intelligence **228**, 195–216 (2015)
13. Euzenat, J.: Interaction-based ontology alignment repair with expansion and relaxation. In: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017, pp. 185–191 (2017). DOI 10.24963/ijcai.2017/27
14. Euzenat, J., Shvaiko, P.: Ontology Matching, Second Edition. Springer (2013)
15. Halpern, J.Y.: Alternative semantics for unawareness. Games and Economic Behavior **37**(2), 321–339 (2001)
16. Jiménez-Ruiz, E., Meilicke, C., Cuenca Grau, B., Horrocks, I.: Evaluating mapping repair systems with large biomedical ontologies. In: Proceedings of the 26th International Workshop on Description Logics, Ulm, Germany, July 23 - 26, 2013, pp. 246–257 (2013). URL `http://ceur-ws.org/Vol-1014/paper_63.pdf`
17. 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, KR 2016, Cape Town, South Africa, April 25-29, 2016., pp. 217–226 (2016). URL `http://www.aaai.org/ocs/index.php/KR/KR16/paper/view/12893`
18. 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 multiagent systems, pp. 1–8 (2007)
19. Meilicke, C.: Alignment incoherence in ontology matching. Ph.D. thesis, University of Mannheim (2011). URL `https://ub-madoc.bib.uni-mannheim.de/29351`

20. Payne, T., Tamma, V.: Negotiating over ontological correspondences with asymmetric and incomplete knowledge. In: International conference on Autonomous Agents and Multi-Agent Systems, AAMAS '14, Paris, France, May 5-9, 2014, pp. 517–524 (2014). URL http://dl.acm.org/citation.cfm?id=2615816
21. 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)
22. Santos, E., Faria, D., Pesquita, C., Couto, F.: Ontology alignment repair through modularization and confidence-based heuristics. Tech. rep. (2013). URL http://arxiv.org/abs/1307.5322
23. Shvaiko, P., Euzenat, J.: Ontology matching: State of the art and future challenges. IEEE Trans. Knowl. Data Eng. **25**(1), 158–176 (2013). DOI 10.1109/TKDE.2011.253
24. Steels, L. (ed.): Experiments in cultural language evolution, vol. 3. John Benjamins Publishing (2012)
25. Tamma, V., Cranefield, S., Finin, T., Willmott, S. (eds.): Ontologies for agents. Birkhäuser (2005)
26. 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)
27. 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
28. Van Benthem, J.: Logical dynamics of information and interaction. Cambridge University Press (2011)
29. Van Benthem, J., Van Eijck, J., Kooi, B.: Logics of communication and change. Inf. Comput. **204**(11), 1620–1662 (2006). DOI 10.1016/j.ic.2006.04.006
30. Van Benthem, J., Velázquez-Quesada, F.: The dynamics of awareness. Synthese **177**(1), 5–27 (2010)
31. 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)
32. 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)
33. Van Ditmarsch, H., Herzig, A., Lang, J., Marquis, P.: Introspective forgetting. Synthese **169**(2), 405–423 (2009)
34. Van Ditmarsch, H., Van der Hoek, W., Kooi, B.: Dynamic epistemic logic, vol. 337. Springer Science & Business Media (2007)