Journal of Artificial Intelligence Research 49 (2014) 363-402 Submitted 09/13; published 02/14

Multiagent Only Knowing in Dynamic Systems

Vaishak Belle VAISHAK@CS.TORONTO.EDU
Dept. of Computer Science, University of Toronto,
Toronto, Ontario, Canada M5S 3H5

Gerhard Lakemeyer GERHARD@CS.RWTH-AACHEN.DE
Dept. of Computer Science, RWTH Aachen University,
52056 Aachen, Germany

Abstract

The idea of "only knowing" a collection of sentences, as proposed by Levesque, has been previously shown to be very useful in characterizing knowledge-based agents: in terms of a specification, a precise and perspicuous account of the beliefs and non-beliefs is obtained in a monotonic setting. Levesque's logic is based on a first-order modal language with quantifying-in, thus allowing for *de re* versus *de dicto* distinctions, among other things. However, the logic and its recent dynamic extension only deal with the case of a single agent. In this work, we propose a first-order multiagent framework with knowledge, actions, sensing and only knowing, that is shown to inherit all the features of the single agent version. Most significantly, we prove reduction theorems by means of which reasoning about knowledge and actions in the framework simplifies to non-epistemic, non-dynamic reasoning about the initial situation.

1. Introduction

When considering knowledge-based agents in dynamic worlds, much depends on what is known and what is not, and how that evolves. Making a telephone call, for example, requires knowing the referent, and if the number is not known, a lookup in the telephone directory must be attempted, after which the agent would have sufficient information to complete the task. Essentially, the agent deliberates on the act of sensing when the agent's knowledge base (KB) informs the agent that it is ignorant of some fact, perhaps one that is necessary for achieving a goal. Moreover, taking a pragmatic point of view, it is desirable that when designing agents, the modeler would provide certain facts but may leave others unsaid. Think of a simple blocks world domain where we find a red block on the table. The agent would be told of the red block, but in the absence of a complete description of the location of every other block in the domain, the agent has to make do with partial information. Thus, at the very least, what is needed is a compact way to write down the knowledge base, thereby providing a full specification of the beliefs and non-beliefs, an account of how that changes after acting and sensing, and a query language that explicitly refers to this knowledge.[1]

One way to view the first requirement is to think that the beliefs of the agent are exactly those that follow from the assumption that a KB is *all* that is believed. Perhaps the most general account to capture the beliefs of a KB is $\mathcal{OL}$: the logic of *only knowing* by Levesque (1990). Levesque's proposal is remarkably simple. He augments a logic of belief (Hintikka, 1962; Kripke, 1963; Fagin,

1. We use the terms "knowledge" and "belief" interchangeably with the understanding that knowledge need not necessarily be true in the real world.

Halpern, Moses, & Vardi, 1995), where (say) the modality $\boldsymbol{K}$ denotes knowledge, with a modality $\boldsymbol{O}$ to capture the notion of "only knowing." Beliefs are reasoned about in terms of valid sentences of the form:

$$\boldsymbol{O}\text{KB} \supset \boldsymbol{K}\alpha$$

which is to be read as "if KB is all that is believed by the agent, then the agent knows α." What is particularly interesting about the new modality is that it not only allows one to draw conclusions about what is known but also about what is not. That is, $\boldsymbol{O}p \supset \neg\boldsymbol{K}q$ and, by introspection, $\boldsymbol{O}p \supset \boldsymbol{K}\neg\boldsymbol{K}q$ both come out valid. Note that this is quite different from classical epistemic logic (Fagin et al., 1995), in the sense that if we replace $\boldsymbol{O}$ by $\boldsymbol{K}$, then neither of these sentences is valid. As a consequence, for example, from $\boldsymbol{O}(Tel(A, 1234) \vee Tel(B, 1234))$ the agent concludes $\boldsymbol{K}(\exists x.\ Tel(x, 1234) \wedge \neg\boldsymbol{K}Tel(x, 1234))$. This says that the agent knows there is someone whose telephone number is *1234* without knowing who, usually referred to as the *de dicto* versus *de re* distinctions in knowledge (Kaplan, 1968). Thus, an agent is able to reason about its own ignorance, in a quantificational language, without having to be told explicitly what it does not know.

While $\mathcal{OL}$ does capture the desiderata on beliefs, it does not include any notions of actions. To obtain the many features of $\mathcal{OL}$ in a dynamic setting, a logic $\mathcal{ES}$ (Lakemeyer & Levesque, 2011) was proposed that amalgamates $\mathcal{OL}$ and the situation calculus (McCarthy & Hayes, 1969; Reiter, 2001; Scherl & Levesque, 2003). The situation calculus is a popular and general formalism for representing and reasoning about dynamic domains. $\mathcal{ES}$ is a (situation-suppressed) modal dialect of the situation calculus[2] that has formulas like those of traditional dynamic logic (Harel, Kozen, & Tiuryn, 2000), such as

$$[pickup(obj5)](Holding(obj5) \wedge \neg Holding(obj3))$$

which says that after picking up *obj5*, the agent is holding *obj5* but not *obj3*. In $\mathcal{ES}$, one stipulates the set of axioms capturing the application domain to be all that is known by the agent, and then obtains entailments regarding beliefs, non-beliefs, and belief *expansion* that can resolve the agent's ignorance as it acts and perceives in the environment. For example,

$$\phi = \{\Box SF(senseFragility(x)) \equiv Fragile(x)\},$$

roughly says that after any sequence of actions, if the agent were to perform a fragility sensing action, *SF* would inform the robot whether the object sensed is fragile or not. If *obj5* is an object that is fragile in the real world, $\mathcal{ES}$ allows us to reason about entailments of the sort:

1. $\models \phi \wedge Fragile(obj5) \wedge \boldsymbol{O}\phi \supset \neg\boldsymbol{K}(Fragile(obj5))$;

2. $\models \phi \wedge Fragile(obj5) \wedge \boldsymbol{O}\phi \supset [senseFragility(obj5)]\boldsymbol{K}(Fragile(obj5))$;

which, in English, says that although the agent does not know that *obj5* is fragile initially, he does so after sensing.

$\mathcal{ES}$ not only allows Reiter-style basic action theories, but is also equipped with an important result from (Reiter, 2001; Scherl & Levesque, 2003): the regression theorem for knowledge. That

2. Under certain assumptions, valid sentences in $\mathcal{ES}$ can be mapped as valid sentences in the classical situation calculus (Lakemeyer & Levesque, 2011). That is, $\mathcal{ES}$ can serve as a semantic basis for the situation calculus with a more workable model theory.

is, sentences and goals about the future, even those mentioning belief, are reduced to questions (perhaps involving knowledge) about the initial state only. More importantly, a significant result from $\mathcal{OL}$ called the *representation theorem* (Levesque & Lakemeyer, 2001) can be leveraged to reduce epistemic queries about the initial state to a first-order reasoning task. In effect, no modal reasoning will be necessary.

However, $\mathcal{ES}$ only deals with the single agent case. Many AI applications where such formalisms are needed involve multiple agents. We might imagine a robot following the lead of another agent, perhaps a second robot, and they are to coordinate deliveries of items between rooms. Similarly, we imagine two agents playing a game of cards against each other. In these and in others, modeling and reasoning about beliefs and non-beliefs that agents have about the real world and the other agents in this world is of interest. In the case of a card game, for example, especially a fair one, agents might believe initially that all their opponents know are the rules of the game. This might then justify certain strategies that depend on the lack of information on the opponent's part.

Before extending $\mathcal{ES}$ to the multiagent case, however, we first need an account of only knowing in the multiagent case. While a number of previous proposals (Lakemeyer, 1993; Halpern, 1993; Halpern & Lakemeyer, 2001; Waaler & Solhaug, 2005) have attempted multiagent extensions to $\mathcal{OL}$, they are all propositional. Besides, they significantly deviate from Levesque's simple model theory. In recent work (Belle & Lakemeyer, 2010a), we were able to show that a natural generalization of $\mathcal{OL}$ to the *n*-agent case does exist for a first-order language. In this article, we continue that line of work and propose a *n*-agent generalization to $\mathcal{ES}$.[3] For the *projection* problem (Reiter, 2001), where we are interested in reasoning about goals (perhaps involving multiagent beliefs) after actions, we show that a regression property is provable. Finally, we also obtain a representation theorem for the *n*-agent case by means of which no modal reasoning will be necessary. We survey related literature in greater detail in Section 5 but both of these results differ from existing results in the epistemic situation calculus (Scherl & Levesque, 2003), which extends the situation calculus in having a notion of knowledge realized in terms of an accessibility relation between situations. (That is, situations are viewed as possible worlds.) For instance, the regression property is different from previous multiagent generalizations (Shapiro, Lespérance, & Levesque, 2002; Kelly & Pearce, 2008) of the epistemic situation calculus in that the background theory may involve nesting of only knowing operators, such as "all that Alice knows is that Bob only knows the rules of the game." Capturing multiagent only knowing in possible-world models that include explicit accessibility relations between worlds (Fagin et al., 1995), as required by the classical epistemic situation calculus, is known to be problematic (Halpern & Lakemeyer, 2001; Belle & Lakemeyer, 2010a), and so such statements do not have obvious counterparts in previous proposals. Similarly, the reduction of knowledge to first-order reasoning is investigated in a very restricted setting by Reiter (2001), and for the single agent case only.

The paper is structured as follows. We first introduce the logic, followed by a discussion of basic action theories. Subsequently, we prove the regression property and a generalized representation theorem. We end after discussing related work. Appendices contain proofs of the main results, that is, the regression property and the representation theorem.

3. A preliminary version of this work appears in the proceedings of the Twenty-Fourth AAAI Conference on Artificial Intelligence, Atlanta, Georgia, USA, July 11-15, 2010 (Belle & Lakemeyer, 2010b).

2. The Formalism

We let $\mathcal{ES}_n$ be a first-order modal language consisting of formulas over symbols from the following vocabulary:

- first-order variables of the object sort: $x_1, x_2, \ldots, y_1, y_2, \ldots$;
- first-order variables of the action sort: $a_1, a_2, \ldots$;
- fluent predicates of arity k: $F_1, F_2, \ldots$; for example, *Wet*;
- rigid predicates of arity k: $G_1, G_2, \ldots$; for example, *Fragile*;
- fluent function symbols of arity k: $f_1, f_2, \ldots$; for example, *distance*;
- rigid function symbols of arity k: $g_1, g_2, \ldots$; for example, *pickup, senseColor*;
- countably infinite standard names: $^{\#}1, ^{\#}2, \ldots$ for objects and actions;
- connectives and other symbols: $=, \vee, \neg, \forall, \boldsymbol{K}_i, \boldsymbol{O}_i, [a], \square$, parenthesis, period and comma.

In the following, for ease of exposition, we assume $i \in \{A, B\}$ in $\boldsymbol{K}_i$ and $\boldsymbol{O}_i$, that is, there are two agents A and B. The extension to more agents is straightforward.

We remark that standard names are *rigid designators*, that is, they mean the same entity in all possible worlds (see below). They can be thought of as constants but satisfying the unique name assumption and an infinitary version of domain closure. Having these symbols means that quantification can be understood substitutionally. Readers familiar with the classical situation calculus (Reiter, 2001) may note that situation terms do not appear in the language. Therefore, we have to distinguish fluents, whose values change after actions, and rigids, whose values do not, both syntactically as well as semantically. $\mathcal{ES}_n$ is also assumed to contain a distinguished predicate *Poss* and distinguished functions SF_i, both of which take an action as an argument. Essentially, *Poss*(a) says that a is executable; $SF_i(a)$ refers to agent i's sensing outcomes on performing a, as shown for the single agent case in the previous section using the fragility sensing action. Section 3 will discuss this in detail for multiple agents.

The *terms* of $\mathcal{ES}_n$ are of the sort action or object, and they are the least set such that:

- every standard name and first-order variable is a term of the corresponding sort;
- if $t_1, \ldots, t_k$ are terms (of any sort) and f is a k-ary function, then $f(t_1, \ldots, t_k)$ is a term.

By a *primitive term*, we mean one of the form $f(n_1, \ldots, n_k)$ where f is a (fluent or rigid) function symbol and all of the n_i are standard names.

The *well-formed formulas* of $\mathcal{ES}_n$ form the least set such that:

- if $t_1, \ldots, t_k$ are terms, and F is a k-ary predicate symbol then $F(t_1, \ldots, t_k)$ is an (atomic) formula;
- if t_1 and t_2 are terms, then $(t_1 = t_2)$ is a formula;

- if t is an action term and α is a formula then $[t]\alpha$ is a formula;
- if α and β are formulas, and x is a first-order variable then the following are also formulas: $\neg\alpha, \alpha \vee \beta, \forall x\alpha, \Box\alpha, \boldsymbol{K}_i\alpha, \boldsymbol{O}_i\alpha$.

  As usual, we treat other connectives such as $\supset$ and $\equiv$ as abbreviations. That is, $\alpha \supset \beta$ abbreviates $\neg\alpha \vee \beta$, and $\alpha \equiv \beta$ abbreviates $(\alpha \supset \beta) \wedge (\beta \supset \alpha)$.

$\mathcal{ES}_n$ has two epistemic modalities. We read $\boldsymbol{K}_i\alpha$ as "i knows α," and we read $\boldsymbol{O}_i\alpha$ as "all that i knows is α." $\mathcal{ES}_n$ also includes dynamic modalities. We read $[a]\alpha$ as "α holds after doing a" and we read $\Box\alpha$ as "α holds after all possible action *sequences*."

A formula without any free variables is called a *sentence*. We also refer to certain kinds of formulas with the following terminology:

- A formula with no $\Box$ operators is called *bounded*.
- A formula with no $[t]$ or $\Box$ operators is called *static*.
- A formula that does not mention $\boldsymbol{O}_i$ for any i is called *basic*. (The formula may mention $\boldsymbol{K}_A$ or $\boldsymbol{K}_B$.)
- A formula with no $\boldsymbol{K}_i, \boldsymbol{O}_i, [t], Poss$ or SF_i is called *fluent*.[4]

For example, $P(^\#1) \vee [t]\boldsymbol{K}_A P(^\#2)$ is bounded, but not static; $P(^\#1) \vee \boldsymbol{K}_A P(^\#2)$ is static and basic, but it is not a fluent formula; $P(^\#1) \vee \boldsymbol{O}_A P(^\#2) \vee Poss(t)$ is a static formula, but neither is it a basic formula nor is it a fluent formula; $(P(^\#1) \vee P(^\#2)) \wedge Q(f(^\#3))$ is a fluent formula.

2.1 The Semantics

A semantics is provided in terms of *possible worlds*. The purpose of the semantics is to determine the values of fluents, both initially and after any sequence of actions. Therefore, in $\mathcal{ES}_n$, similar to the idea of situation trees (Reiter, 2001), worlds determine the changing values of fluents after actions; see Figure 1 for the intuition. More precisely,

- let $\mathcal{Z}$ denote all finite sequences of action names, including $\langle\rangle$, which is the empty sequence (corresponding to the initial situation);
- then a world $w \in \mathcal{W}$ is any function from $\mathcal{G} \times \mathcal{Z}$ to $\{0, 1\}$, where $\mathcal{G}$ is the set of primitive atoms, and from $\mathcal{T} \times \mathcal{Z}$ to $\mathcal{N}$ (preserving sorts), where $\mathcal{T}$ is the set of primitive terms, and satisfying the rigidity constraint: if g is a rigid function or predicate symbol, then for all z and z' in $\mathcal{Z}$, $w[g(n_1, ..., n_k), z] = w[g(n_1, ..., n_k), z']$.

To interpret arbitrary terms, we proceed as follows. As mentioned earlier, names are rigid designators. Given a term t without variables, a world w and a sequence z, we define $|t|_w^z$ (to be read as "the co-referring standard name for t given w and z") by:

1. $|t|_w^z = t$ if t is a name;

4. In the situation calculus (Reiter, 2001), these correspond to formulas that are *uniform* in a situation term.

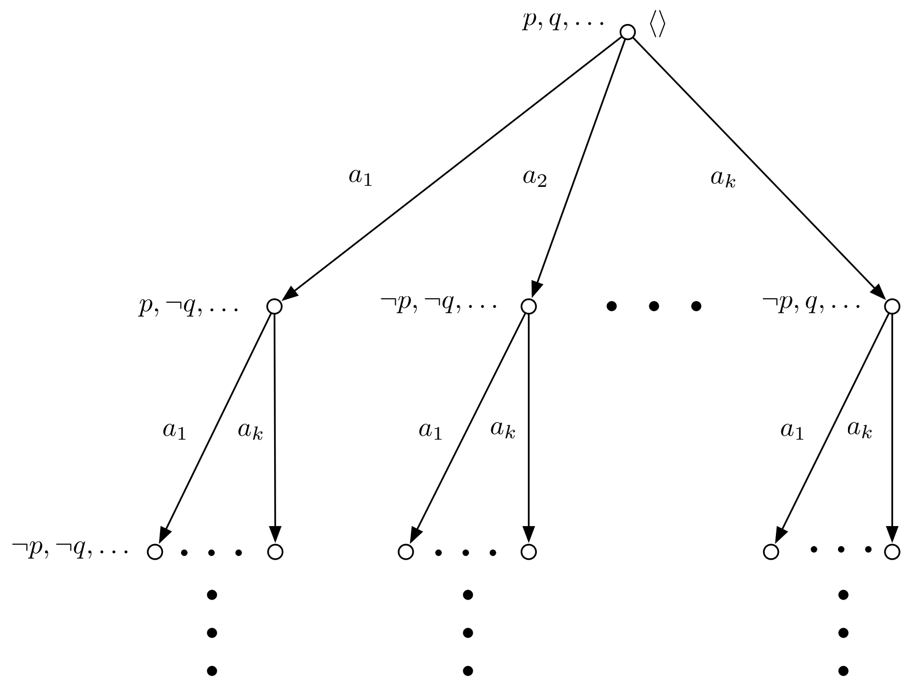

Figure 1: A possible world.

2. $|f(t_1, \ldots, t_k)|^z_w = w[f(n_1, \ldots, n_k), z]$, where $|t_i|^z_w = n_i$.

Agents may, of course, have incomplete knowledge. To distinguish their uncertainty from the real world, we stipulate *epistemic states* that model multiple possibilities. Standard accounts of multiagent epistemic states are based on Kripke frames (Fagin et al., 1995). For multiagent only knowing, however, Kripke-based accounts turn out to be very problematic, as seen in the work of Halpern (1993), Lakemeyer (1993), Halpern and Lakemeyer (2001) and Waaler and Solhaug (2005). For example, Lakemeyer (1993) shows that certain types of epistemic states cannot be constructed in his approach. In the work of Halpern (1993), epistemic operators do not interact in an intuitive manner (Halpern & Lakemeyer, 2001). In the work of Halpern and Lakemeyer (2001), the semantic notion of validity is defined directly in the language, making the proposal unnatural. Serious complications are present in later proposals as well (Waaler & Solhaug, 2005). Moreover, none of these have been extended to a quantified language. A discussion on these issues is not needed for the purposes of this article; interested readers are referred to our earlier work (Belle & Lakemeyer, 2010a). In that work, we then proposed an alternative called k-structures, which was shown to generalize Levesque's (1990) proposal to the many agent case in an appropriate and intuitive manner. These structures deviate from Kripke-based accounts in defining epistemic states of increasing depths. As it turns out, these structures also have a natural extension to the dynamic setting, which we present below. We first define a notion of *depth* for formulas in the following way:

Definition 1 The i-depth of $\alpha \in \mathcal{ES}_n$, denoted $|\alpha|_i$, is defined inductively as ($\boldsymbol{M}_i$ denotes $\boldsymbol{K}_i$ or $\boldsymbol{O}_i$):

- $|\alpha|_i = 1$ for atomic formulas;
- $|\neg\alpha|_i = |\alpha|_i$;

- $|\forall x\alpha|_i = |\alpha|_i$;
- $|[a]\alpha|_i = |\alpha|_i$;
- $|\Box\alpha|_i = |\alpha|_i$;
- $|\alpha \vee \beta|_i = max(|\alpha|_i, |\beta|_i)$;
- $|\boldsymbol{M}_i\alpha|_i = |\alpha|_i$,
- $|\boldsymbol{M}_j\alpha|_i = |\alpha|_j + 1$ for $j \neq i$.

A formula α has depth k if $max(|\alpha|_A, |\alpha|_B) = k$. ∎

Given a formula of A-depth k and of B-depth j, we say that the formula has A, B-depth of k, j for brevity. We say α is *objective* if no epistemic operators are mentioned in α. A formula is called *i-objective* if all epistemic operators which do not occur within the scope of another epistemic operator are of the form $\boldsymbol{M}_j$, $j \neq i$, where $\boldsymbol{M}_i$ denotes $\boldsymbol{K}_i$ or $\boldsymbol{O}_i$. A formula is called *i-subjective* if every atom is in the scope of an epistemic operator and all epistemic operators which do not occur within the scope of another epistemic operator are of the form $\boldsymbol{M}_i$. Intuitively, i-subjective formulas represent i's beliefs about the world whereas i-objective formulas determine what is true about the world from i's perspective, which may include beliefs of agents other than i.

Example 2 Consider the formula $\Box\boldsymbol{K}_A\boldsymbol{K}_B\boldsymbol{K}_A p \vee \boldsymbol{K}_B[t]q$. Here:

- $|\Box\boldsymbol{K}_A\boldsymbol{K}_B\boldsymbol{K}_A p \vee \boldsymbol{K}_B[t]q|_A = max(|\Box\boldsymbol{K}_A\boldsymbol{K}_B\boldsymbol{K}_A p|_A, |\boldsymbol{K}_B[t]q|_A) = 3$ because
  1. $|\Box\boldsymbol{K}_A\boldsymbol{K}_B\boldsymbol{K}_A p|_A = |\boldsymbol{K}_A\boldsymbol{K}_B\boldsymbol{K}_A p|_A = |\boldsymbol{K}_B\boldsymbol{K}_A p|_A = 1 + |\boldsymbol{K}_A p|_B = 2 + |p|_A = 3$,
  2. $|\boldsymbol{K}_B[t]q|_A = 1 + |[t]q|_B = 1 + |q|_B = 2$.
- $|\Box\boldsymbol{K}_A\boldsymbol{K}_B\boldsymbol{K}_A p \vee \boldsymbol{K}_B[t]q|_B = max(|\Box\boldsymbol{K}_A\boldsymbol{K}_B\boldsymbol{K}_A p|_B, |\boldsymbol{K}_B[t]q|_B) = 4$ because
  1. $|\Box\boldsymbol{K}_A\boldsymbol{K}_B\boldsymbol{K}_A p|_B = |\boldsymbol{K}_A\boldsymbol{K}_B\boldsymbol{K}_A p|_B = 1 + |\boldsymbol{K}_B\boldsymbol{K}_A p|_A = 1 + 3$ (as shown above) $= 4$,
  2. $|\boldsymbol{K}_B[t]q|_B = |[t]q|_B = |q|_B = 1$.

Therefore, the depth of the formula is 4. Consider each of the disjuncts. $\Box\boldsymbol{K}_A\boldsymbol{K}_B\boldsymbol{K}_A p$ is both A-subjective as well as B-objective. On the other hand, $\boldsymbol{K}_B[t]q$ is both B-subjective as well as A-objective. Moreover, $\Box\boldsymbol{K}_A\boldsymbol{K}_B\boldsymbol{K}_A p \vee \boldsymbol{K}_B[t]q$ is neither A-subjective nor B-subjective. For that matter, it is neither A-objective nor B-objective. ∎

The beliefs of an agent are captured by means of a k-structure defined over the set $\mathcal{W}$:

Definition 3 A k-structure e^k, where $k \geq 1$, is defined inductively as:

- $e^1 \subseteq \mathcal{W} \times \{\{\}\}$,
- $e^k \subseteq \mathcal{W} \times \mathbb{E}^{k-1}$, where $\mathbb{E}^m$ is the set of all m-structures. ∎

That is, a e^1 is simply a set of worlds. A e^2 is a set of the form $\{(w, e^1), (w', e'^1), \ldots\}$ which states that at w an agent, say A, believes B to consider worlds from e^1 possible, and at w' she believes B to consider worlds from e'^1 possible. This captures the intuition that A has partial information about B, and so her beliefs about B differ at different worlds.[5] When modeling a k-structure, say e^k, for A we denote it as e^k_A. Analogously, when modeling a j-structure, say e^j, for B we denote it as e^j_B.

Such structures essentially represents the initial beliefs of the agent, that is, the initial state of knowledge. But when actions occur, perhaps an agent acquires new information and as a result of this some of the possibilities in an epistemic state may be discarded over the course of doing actions (Scherl & Levesque, 2003). Following Lakemeyer and Levesque (2011), we capture this feature by means of a *compatibility relation* $\simeq^i_z$ between worlds (relative to an agent i), which looks for truth in the real world by means of sensing. We define $w' \simeq^i_z w$ inductively by the following:

- $w' \simeq^i_{\langle\rangle} w$ for all worlds w' and w;
- $w' \simeq^i_{z\cdot r} w$ iff $w' \simeq^i_z w$ and $w'[SF_i(r), z] = w[SF_i(r), z]$.

We define a e^k for A, a e^j for B and a world w as a (k, j)-model (e^k_A, e^j_B, w). The idea is that *only* formulas with a maximal A-depth of k and with a maximal B-depth of j are to be interpreted wrt (k, j)-models. To determine whether a formula is true or not after a sequence of actions z given a (k, j)-model, we write $e^k_A, e^j_B, w, z \models \alpha$. The definition of truth is as follows:

1. $e^k_A, e^j_B, w, z \models P(t_1, \ldots, t_k)$ iff $w[P(n_1, \ldots, n_k), z] = 1$ where $|t_i|^z_w = n_i$;
2. $e^k_A, e^j_B, w, z \models t_1 = t_2$ iff n_1 and n_2 are the same standard names, where $|t_i|^z_w = n_i$;
3. $e^k_A, e^j_B, w, z \models \neg\alpha$ iff $e^k_A, e^j_B, w, z \not\models \alpha$;
4. $e^k_A, e^j_B, w, z \models \alpha \vee \beta$ iff $e^k_A, e^j_B, w, z \models \alpha$ or $e^k_A, e^j_B, w, z \models \beta$;
5. $e^k_A, e^j_B, w, z \models \forall x\alpha$ iff $e^k_A, e^j_B, w, z \models \alpha^x_n$ for every name n of the appropriate sort;
6. $e^k_A, e^j_B, w, z \models [t]\alpha$ iff $e^k_A, e^j_B, w, z \cdot r \models \alpha$ where $|t|^z_w = r$;
7. $e^k_A, e^j_B, w, z \models \Box\alpha$ iff $e^k_A, e^j_B, w, z \cdot z' \models \alpha$ for every $z' \in \mathcal{Z}$;
8. $e^k_A, e^j_B, w, z \models \boldsymbol{K}_A\alpha$ iff for all $w' \simeq^A_z w$, for all e^{k-1} (for B),
   if $(w', e^{k-1}_B) \in e^k_A$ then $e^k_A, e^{k-1}_B, w', z \models \alpha$;
9. $e^k_A, e^j_B, w, z \models \boldsymbol{O}_A\alpha$ iff for all $w' \simeq^A_z w$, for all e^{k-1} (for B),
   $(w', e^{k-1}_B) \in e^k_A$ iff $e^k_A, e^{k-1}_B, w', z \models \alpha$.

In an analogous fashion, the semantics for $\boldsymbol{K}_B\alpha$ and $\boldsymbol{O}_B\alpha$ are specified. Here, $\boldsymbol{K}_i$ is the classical epistemic operator. We may read $\boldsymbol{K}_i\alpha$ as "(at least) α is believed" because $\boldsymbol{K}_i\alpha$ certainly does not preclude $\boldsymbol{K}_i(\alpha \wedge \beta)$ from holding in general. On the other hand, if $\boldsymbol{O}_i\alpha$ holds then the epistemic

5. Levesque's (1990) notion of an epistemic state is simply a *set* of worlds. It is easy to see that if there is only a single agent then we only need 1-structures, which then coincides with Levesque's account.

state is one which contains all and only the structures satisfying α. In essence (Levesque, 1990), the definition for $\boldsymbol{O}_i$ differs from that for $\boldsymbol{K}_i$ in using an "iff" rather than an "if".[6]

Given a sentence α of maximal A, B-depth k, j, we write $e^k_A, e^j_B, w \models \alpha$ to mean $e^k_A, e^j_B, w, \langle\rangle \models \alpha$. We say that a sentence α of maximal A, B-depth k, j is satisfiable if there is a (k, j)-model (e^k_A, e^j_B, w) such that $e^k_A, e^j_B, w \models \alpha$. If Σ is any set of sentences of maximal A, B-depth of k, j and α is as above, we write $\Sigma \models \alpha$ (read: "Σ entails α") iff for every (k, j)-model such that $e^k_A, e^j_B, w \models \alpha'$ for every $\alpha' \in \Sigma$ then $e^k_A, e^j_B, w \models \alpha$. We write $\models \alpha$ (read: "α is valid") to mean $\{\} \models \alpha$.

We often write $\{\}, e^j_B, w \models \alpha$ when α is A-objective because the k-structure for A is irrelevant. Analogously, for B-objective formulas, we often write $e^k_A, \{\}, w \models \alpha$. When the formula α is objective, we omit the structures for A and B altogether and simply write $w \models \alpha$.

2.2 Properties

We differ slightly from usual semantical accounts in that the satisfaction relation is undefined for formulas whose depth exceeds a certain number. Nevertheless, we are able to show that as far as entailment is concerned, such an account does not present any serious limitations. Let us begin with a few simple examples.

Example 4 Let p be an atom. Then the following sentences are valid. Our method for proving these examples will be to look at the sentence to decide on the depth of the models. (We will use TRUE to denote a tautologous sentence, such as $\forall x.\ (x = x)$.)

1. $\boldsymbol{O}_A\text{TRUE} \supset \neg\boldsymbol{K}_A\neg\boldsymbol{K}_B p$.

   The sentence is A-subjective and of A-depth 2. So consider any 2-structure for A that satisfies $\boldsymbol{O}_A$TRUE. Here is one: let $e^2_A = \mathcal{W} \times 2^{\mathcal{W}}$. Clearly $e^2_A, \{\}, w \models \boldsymbol{O}_A\text{TRUE}$. (We reiterate that when the epistemic state for B is irrelevant, we will simply write $(e^k_A, \{\}, w)$ and ignore the structure for B.) It is easy to verify that no other e^2 satisfies $\boldsymbol{O}_A$TRUE. So now $e^2_A, \{\}, w \models \neg\boldsymbol{K}_A\neg\boldsymbol{K}_B p$ iff there is some $(w', e^1_B) \in e^2_A$ such that $e^2_A, e^1_B, w' \models \boldsymbol{K}_B p$. By construction, there is $(w, e^{*1}_B) \in e^2_A$ where $e^{*1}_B = \{(w, \{\}) \mid w \models p\}$ and $e^2_A, e^{*1}_B, w \models \boldsymbol{K}_B p$.

2. $\boldsymbol{O}_A\text{TRUE} \supset \neg\boldsymbol{K}_A\boldsymbol{K}_B p$.

   Construct e^2_A as in item 1. Then $e^2_A, \{\}, w \models \neg\boldsymbol{K}_A\boldsymbol{K}_B p$ iff there is some $(w', e^1_B) \in e^2_A$, such that $e^2_A, e^1_B, w' \models \neg\boldsymbol{K}_B p$. By construction, $(w, e^{*1}_B) \in e^2_A$ where $e^{*1}_B = \{(w, \{\}) \mid w \not\models p\}$ and, $e^2_A, e^{*1}_B, w \models \neg\boldsymbol{K}_B p$.

3. $\boldsymbol{O}_A(p \wedge \boldsymbol{O}_B p) \supset \boldsymbol{K}_A p$.

   We will consider any 2-structure for A satisfying $\boldsymbol{O}_A(p \wedge \boldsymbol{O}_B p)$ and prove that $\boldsymbol{K}_A p$ is also satisfied at the structure. So let $\mathcal{W}_p = \{w \mid w \models p\}$. Clearly $e^1_B = \{(w, \{\}) \mid w \in \mathcal{W}_p\}$ is the only 1-structure for B that satisfies $\boldsymbol{O}_B p$. Similarly, the 2-structure $e^2_A = \{(w, e^1_B) \mid w \in \mathcal{W}_p\}$ is the only 2-structure for A that satisfies $\boldsymbol{O}_A(p \wedge \boldsymbol{O}_B p)$. It follows that $e^2_A, \{\}, w \models \boldsymbol{K}_A p$ since all w' in $(w', e^1_B) \in e^2_A$ satisfy p by construction.

6. In the literature, for a closer examination of the relationship between these modalities, a third modality to denote what the agent "at most" knows is often included in the logical language (Halpern & Lakemeyer, 2001; Levesque & Lakemeyer, 2001). This modality need not concern us here. We refer interested readers to our earlier work on how a semantics is given for such an operator using k-structures (Belle & Lakemeyer, 2010a).

4. $\boldsymbol{O}_A(p \wedge \boldsymbol{O}_B p) \supset \boldsymbol{K}_A \boldsymbol{K}_B p$.

   A 2-structure e_A^2 is constructed as in item 3. Then it follows that $e_A^2, \{\}, w \models \boldsymbol{K}_A \boldsymbol{K}_B p$ since all worlds
   $$\{w'' \mid (w'', \{\}) \in e_B^1 \text{ and } (w', e_B^1) \in e_A^2 \text{ for some } w'\}$$
   satisfy p by construction.

5. $\boldsymbol{O}_A(p \wedge \boldsymbol{O}_B p) \supset (\boldsymbol{K}_A \neg \boldsymbol{K}_B \boldsymbol{K}_A p \wedge \boldsymbol{K}_A \neg \boldsymbol{K}_B \neg \boldsymbol{K}_A p)$.

   Using ideas from item 1 and 2, it follows that $\boldsymbol{O}_B p \supset \neg \boldsymbol{K}_B \boldsymbol{K}_A p \wedge \neg \boldsymbol{K}_B \neg \boldsymbol{K}_A p$ is valid. Let e_A^3 be any structure that satisfies $\boldsymbol{O}_A(p \wedge \boldsymbol{O}_B p)$. Since for all $(w', e_B^2) \in e_A^3$, $e_A^3, e_B^2, w' \models p \wedge \boldsymbol{O}_B p$, it follows that $e_A^3, e_B^2, w' \models \neg \boldsymbol{K}_B \boldsymbol{K}_A p \wedge \neg \boldsymbol{K}_B \neg \boldsymbol{K}_A p$. Therefore $e_A^3, \{\}, w \models \boldsymbol{K}_A(\neg \boldsymbol{K}_B \boldsymbol{K}_A p \wedge \neg \boldsymbol{K}_B \neg \boldsymbol{K}_A p)$.

Items 1 and 2 tell us if all that A knows is TRUE, then she correctly reasons about her ignorance: she does not know *whether* B knows p. Items 3 and 4 tell us if A only knows $\{p, \boldsymbol{O}_B p\}$, then she correctly believes that both she and B believe p. Finally, item 5 tells us since A believes B only knows p, she believes B cannot tell whether A knows p. ∎

In these examples, we (appropriately) chose structures of a certain depth to interpret the sentences of a corresponding depth. However, as far as validity goes, models of any higher depth can be considered. That is, if a formula of maximal A, B-depth k, j is true at all (k, j)-models, then the formula is also true at all (k', j')-models, for $k' \geq k$ and $j' \geq j$. To demonstrate this property, we construct for every $e_A^{k'}$, a k-structure $e_A{\downarrow}_k^{k'}$, such that they agree on all formulas of maximal A-depth k. Analogously, a j-structure that agrees on all formulas of maximal B-depth j can be constructed for every $e_B^{j'}$.

Definition 5 Given $e_A^{k'}$ and $e_B^{j'}$, we inductively define a k-structure $e_A{\downarrow}_k^{k'}$ and a j-structure $e_B{\downarrow}_j^{j'}$ for $k' \geq k \geq 1$ and $j' \geq j \geq 1$, respectively:

- $e_A{\downarrow}_1^1 = e_A^1$;
- $e_B{\downarrow}_1^1 = e_B^1$;
- $e_A{\downarrow}_1^{k'} = \{(w, \{\}) \mid (w, e_B^{k'-1}) \in e_A^{k'}\}$ for $k' > 1$;
- $e_B{\downarrow}_1^{j'} = \{(w, \{\}) \mid (w, e_A^{j'-1}) \in e_B^{j'}\}$ for $j' > 1$;
- $e_A{\downarrow}_k^{k'} = \{(w, e_B{\downarrow}_{k-1}^{k'-1}) \mid (w, e_B^{k'-1}) \in e_A^{k'}\}$ for $k > 1$;
- $e_B{\downarrow}_j^{j'} = \{(w, e_A{\downarrow}_{j-1}^{j'-1}) \mid (w, e_A^{j'-1}) \in e_B^{j'}\}$ for $j > 1$. ∎

With this definition in hand, we get the following property by relating k'-structures $e_A^{k'}$ and j'-structures $e_B^{j'}$, and their corresponding k-structures $e_A{\downarrow}_k^{k'}$ and j-structures $e_B{\downarrow}_j^{j'}$ respectively.

Lemma 6 *Let $k' \geq k$ and $j' \geq j$. For all α of maximum A-depth k and maximum B-depth j:*

$$e_A^{k'}, e_B^{j'}, w \models \alpha \;\; \textit{iff} \;\; e_A{\downarrow}_k^{k'}, e_B{\downarrow}_j^{j'}, w \models \alpha.$$

The proof is not hard, but tedious. The arguments for this result appear elsewhere (Belle & Lakemeyer, 2010a), and so we do not reproduce them here.[7]

Theorem 7 *For all formulas α of A, B-depth of k, j, if α is true at all (k, j)-models, then α is true at all (k', j')-models, where $k' \geq k$ and $j' \geq j$.*

Proof: Suppose α is true at all (k, j)-models. Given any $(e_A^{k'}, e_B^{j'}, w)$, by assumption $e_A{\downarrow}_k^{k'}, e_B{\downarrow}_j^{j'}, w \models \alpha$. By the previous lemma, $e_A^{k'}, e_B^{j'}, w \models \alpha$. ∎

It follows then that one may speak about the valid sentences of the logic without explicitly speculating what their depths or the depths of their models need to be. That is, we may simply assume that models have "appropriate" depths, in the sense of having depths that equal or exceed the depth of the sentences. For example, we obtain the following result that knowledge with k-structures has $\mathbf{K45}_n$ properties (Fagin et al., 1995), as well as the universal and existential versions of the Barcan formula. Moreover, these properties hold after any number of actions have been performed.

Lemma 8 *Let α and β be $\mathcal{ES}_n$-formulas. Then the following sentences are valid:*

1. $\square(\boldsymbol{K}_i\alpha \land \boldsymbol{K}_i(\alpha \supset \beta) \supset \boldsymbol{K}_i\beta)$;
2. $\square(\boldsymbol{K}_i\alpha \supset \boldsymbol{K}_i\boldsymbol{K}_i\alpha)$;
3. $\square(\neg\boldsymbol{K}_i\alpha \supset \boldsymbol{K}_i\neg\boldsymbol{K}_i\alpha)$;
4. $\square(\forall x\boldsymbol{K}_i\alpha \supset \boldsymbol{K}_i\forall x\alpha)$;
5. $\square(\exists x\boldsymbol{K}_i\alpha \supset \boldsymbol{K}_i\exists x\alpha)$.

Proof: The proofs are very similar. We show item 3 and 4. Let i be A. The other case is symmetric.

3. Suppose $e_A^k, e_B^j, w, z \models \neg\boldsymbol{K}_A\alpha$. Then there is some $w' \simeq_z^A w$, $(w', e_B^{k-1}) \in e_A^k$ such that $e_A^k, e_B^{k-1}, w', z \models \neg\alpha$. Let w'' be any world such that $w'' \simeq_z^A w'$, $(w'', e_B'^{\,k-1}) \in e_A^k$. Clearly $e_A^k, e_B'^{\,k-1}, w'', z \models \neg\boldsymbol{K}_A\alpha$. Since $w'' \simeq_z^A w$, we get that $e_A^k, e_B^j, w, z \models \boldsymbol{K}_A\neg\boldsymbol{K}_A\alpha$.
4. Suppose $e_A^k, e_B^j, w, z \models \forall x\boldsymbol{K}_A\alpha$. Then $e_A^k, e_B^j, w, z \models (\boldsymbol{K}_A\alpha)_n^x$ for every name n. That is, $e_A^k, e_B^j, w, z \models \boldsymbol{K}_A\alpha_n^x$ for every n. Then for all $w' \simeq_z^A w$, such that $(w', e_B^{k-1}) \in e_A^k$ we have $e_A^k, e_B^{k-1}, w', z \models \alpha_n^x$ for every n iff by definition $e_A^k, e_B^{k-1}, w', z \models \forall x\alpha$. Therefore $e_A^k, e_B^j, w, z \models \boldsymbol{K}_A\forall x\alpha$. ∎

Apart from $\mathbf{K45}_n$ belief properties, the relationship between only knowing and knowledge can also be established using the notion of validity:

Lemma 9 *Suppose p and q are atoms, and α is any $\mathcal{ES}_n$-formula. Then the following are valid:*

7. While actions are not considered in that work, these are interpreted wrt worlds and so the extension of the argument is straightforward.

1. $\boldsymbol{O}_i\alpha \supset \boldsymbol{K}_i\alpha$;
2. $\boldsymbol{O}_i p \supset \neg\boldsymbol{K}_i q$.

Proof: Item 1 is an easy consequence of the semantics. For item 2, observe that by the definition of only knowing, structures that satisfy $p \wedge \neg q$, which must exist because p and q are atoms, are included in an epistemic state where $\boldsymbol{O}_i p$ holds. Therefore q cannot be known. ∎

Item 1 says that whatever is only known is also believed by the agent. Item 2, of course, relates only knowing and non-beliefs. It is straightforward to generalize the arguments for these properties to also capture the valid sentences from Example 4 involving multiagent nested beliefs.

Finally, when specifying the agent, we want to allow for agents that have false beliefs. This is permitted, and that can be demonstrated by means of the following property that shows that it is possible to know (and only know) a formula that is false in the real world, and it is also possible to not know (and not only know) a formula that is true in the real world.

Lemma 10 *Let p be an atom. Then following sentences are satisfiable (let $\boldsymbol{M}_i$ denote $\boldsymbol{K}_i$ or $\boldsymbol{O}_i$):*

1. $\neg p \wedge \boldsymbol{M}_i p$;
2. $p \wedge \neg\boldsymbol{M}_i p$.

Proof: We show $\boldsymbol{M}_i = \boldsymbol{O}_A$. The case for $\boldsymbol{M}_i = \boldsymbol{O}_B$ is symmetric. The arguments for $\boldsymbol{M}_i = \boldsymbol{K}_i$ is analogous. For item 1, let w be a world such that $w \models \neg p$, and $\mathcal{W}_p = \{w' \mid w' \models p\}$. Let e^1_A be the set $\{(w', \{\}) \mid w' \in \mathcal{W}_p\}$. It follows then that $e^1_A, \{\}, w \models \neg p \wedge \boldsymbol{O}_A p$.

For item 2, suppose $w^* \models p$, and $\mathcal{W}'_p = \mathcal{W}_p - \{w^*\}$. Let $e^1_A = \{(w', \{\}) \mid w' \in \mathcal{W}'_p\}$. Then, we get $e^1_A, \{\}, w^* \models p \wedge \neg\boldsymbol{O}_A p$. ∎

Before concluding this section, let us briefly reflect on the fact that k-structures have a finite depth. So suppose A only knows Σ, of depth k. Using k-structures alone allows us to reason about what is believed and what is not believed, up to depth k. For example, $\boldsymbol{O}_A P(^\#1)$ entails $\boldsymbol{K}_A P(^\#1), \neg\boldsymbol{K}_A P(^\#2), \neg\boldsymbol{K}_A P(^\#3), \ldots$ as shown in Lemma 9. Moreover, as already observed in Example 4, the logic correctly captures that A is ignorant about beliefs at depth greater than k. That is, using the simple example of an agent who only knows TRUE of depth 1, we saw that the sentences $\boldsymbol{O}_A\text{TRUE} \supset \neg\boldsymbol{K}_A\neg\boldsymbol{K}_B p$ and $\boldsymbol{O}_A\text{TRUE} \supset \neg\boldsymbol{K}_A\boldsymbol{K}_B p$ are valid. So, although the KB has finite depth, we are able to ask queries α of any depth in the sense of determining whether the sentence $\boldsymbol{O}_i\Sigma \supset \boldsymbol{K}_i\alpha$ is valid.

For most purposes, this restriction of having a parameter k seems harmless in the sense that agents usually have a finite knowledge base with sentences of some maximal depth k and they should be ignorant about what is known at depths higher than k. But there is one aspect which we cannot handle: the property of simultaneously satisfying an infinite set of sentences of unbounded depth. Indeed, k-structures cannot be used for this purpose simply because, for a fixed k, the satisfaction relation is undefined for formulas beyond depth k.

One prominent application of such a property is the notion of *common knowledge* (Fagin et al., 1995). We do not go over the details here, but the common knowledge modality allows the logic to reason about sentences such as $(\boldsymbol{K}_i\boldsymbol{K}_j)^k\alpha$, where α appears in the scope of k sequences of $\boldsymbol{K}_i\boldsymbol{K}_j$, for any k. Even though the nature of common knowledge is infinitary, in the sense that it essentially corresponds to an infinite conjunction, it can nonetheless be given a finite axiomatic characterization, making it a useful operator for certain applications (Fagin et al., 1995). Thus, if we were to include the notion of common knowledge in a logic, then we would get entailments about what is believed at arbitrary depths. With our current model, however, this cannot be captured. While this is certainly a restriction, we are willing to pay that price because in return we get, for the first time, a very simple model theory for multiagent only knowing (Belle & Lakemeyer, 2010a).

3. Basic Action Theories

Let us now consider the equivalent of basic action theories of the situation calculus. Since situations do not appear in the language, as in $\mathcal{ES}$, the basic action theories do not require foundational axioms like Reiter's second-order induction axiom for situations (Reiter, 2001).

Definition 11 Given a set of fluents $\mathcal{F}$, a set $\Sigma \subseteq \mathcal{ES}_n$ of sentences is called a *basic action theory* (BAT) over $\mathcal{F}$ iff $\Sigma = \Sigma_0 \cup \Sigma_{pre} \cup \Sigma_{post} \cup \Sigma_{sense}$ where Σ only mentions fluents from $\mathcal{F}$ and[8]

1. Σ_0 is any set of fluent sentences;

2. Σ_{pre} is a singleton sentence of the form:

$$\Box Poss(a) \equiv \pi$$

where π is a fluent formula;[9]

3. Σ_{post} is a set that includes sentences of the form:

$$\Box[a]F(\vec{x}) \equiv \gamma_F,$$

one for each fluent predicate F, and sentences of the form:

$$\Box[a]f(\vec{x}) = u \equiv \gamma_f,$$

one for each fluent function f, where γ_F and γ_f are fluent formulas;[10]

4. Σ_{sense} is a set of sentences similar to the one for *Poss* of the form:

$$\Box SF_i(a) = x \equiv \varphi_i,$$

one for each agent i, where φ_i is a fluent formula. ∎

8. We follow the usual convention that free variables are universally quantified from the outside.
9. We assume that $\Box$ has lower syntactic precedence than the logical connectives, so that $\Box Poss(a) \equiv \pi$ stands for $\forall a.\Box(Poss(a) \equiv \pi)$.
10. The $[a]$ construct has higher precedence than the logical connectives. That is, $\Box[a]f(x_1,\ldots,x_k) = y \equiv \gamma_f$ abbreviates $\forall a.\Box([a]f(x_1,\ldots,x_k) = y \equiv \gamma_f)$.

The idea is that Σ_0 expresses what is true initially, Σ_{pre} is one large precondition axiom, Σ_{post} are the successor state axioms, one per fluent, which are formulated so as to incorporate Reiter's solution the frame problem. Σ_{sense} accommodates the intuition that the sensing results for agents may differ for various actions. For example, when B senses that A is reading a letter, we would not expect B to learn the contents of that letter. Here, we follow the convention (Scherl & Levesque, 2003) that every action returns a sensing result. For actions such as *forward*, which do not return any sensing information, SF_i is defined to return a special standard name NIL.

Knowledge about the initial situation may be incomplete. More precisely, we have to distinguish between what is true in the real world and what the agents know or believe about the world. Of course, what A believes about the world may differ from B's knowledge. Moreover, what A believes B to know may differ from what B actually believes. One way to capture such generality is to first insist on an action theory modeling the real world, say Υ. Then, we might imagine differing basic action theories for subsequent levels of beliefs for the agents, as illustrated by the following theory:[11]

$$\Upsilon \wedge \boldsymbol{O}_A(\Sigma \wedge \boldsymbol{O}_B(\Sigma^* \wedge \ldots)) \wedge \boldsymbol{O}_B(\Sigma' \wedge \boldsymbol{O}_A(\Sigma^{**} \wedge \ldots)) \tag{1}$$

where Υ and Σ (with superscripts) are basic action theories that may differ arbitrarily. Here, Υ represents what is true in the real world, and Σ (with superscripts) represent the agent's beliefs. For example, Σ^* represents what A believes B to know. By extension, then, for n agents with k levels, we would expect $n \cdot k + 1$ action theories, each one perhaps differing arbitrarily from each other.

For ease of exposition, we will consider the following simple case in the remainder of the article. The simple case stipulates that if Σ represents A's view of the world, then he believes that Σ also represents B's view of the world. This is reasonable for applications such as simple card games, which we consider below. None of our technical results, including the regression property and the representation theorem, hinge on this stipulation, however. See Section 4.4 for discussions.

A background theory, then, is a special case of (1), as illustrated by the following sentence:

$$\Upsilon \wedge \boldsymbol{O}_A(\Sigma \wedge \boldsymbol{O}_B(\Sigma \wedge \ldots)) \wedge \boldsymbol{O}_B(\Sigma' \wedge \boldsymbol{O}_A(\Sigma' \wedge \ldots)) \tag{2}$$

where, again, Υ, Σ and Σ' may differ arbitrarily.

Formally, in order to prepare for agents that may have beliefs to some arbitrary (but finite) depth, we introduce the following inductive definition over a basic action theory Σ:

- let $OKnow_\Sigma[A, 1] = \boldsymbol{O}_A\Sigma$;
- let $OKnow_\Sigma[B, 1] = \boldsymbol{O}_B\Sigma$;
- for $k > 1$, let $OKnow_\Sigma[A, k] = \boldsymbol{O}_A(\Sigma \wedge OKnow_\Sigma[B, k-1])$;
- for $j > 1$, let $OKnow_\Sigma[B, j] = \boldsymbol{O}_B(\Sigma \wedge OKnow_\Sigma[A, j-1])$.

Given basic action theories Υ, Σ and Σ', in the remainder of the article we will be interested in theories of the form

$$\Upsilon \wedge OKnow_\Sigma[A, k] \wedge OKnow_{\Sigma'}[B, j] \tag{3}$$

11. In the sequel, background theory stipulations assume the nesting of only knowing operators. There are other possibilities, of course, such as $\boldsymbol{O}_i(\phi \wedge (\boldsymbol{K}_j\alpha \vee \boldsymbol{K}_j\beta))$. We defer discussions on these to Section 4.4.

which says that A believes the action theory Σ to k levels, *i.e.* he believes B to also believe Σ and so on, while B believes the action theory Σ' to j levels. Before presenting any technical results on reasoning about actions, we show with an example how the formalism can be used to model domains, and that it has appropriate properties regarding knowledge, introspection and sensing.

Example 12 Imagine two agents playing a simple card game. We imagine a deck of cards, numbered 1 through 52. Two face-down cards have been dealt, one to A and the other to B. Player i picks her card, reads the card and decides to challenge player j ($j \neq i$). When a challenge is posed, the player with the card that has the highest number wins the game.

We begin by stipulating the preconditions of the domain. Let Σ_{pre} be the following:

$$\begin{aligned}&\Box Poss(a) \equiv \\ &\qquad \exists x[a = pick_i(x) \wedge \forall y(\neg Holding_i(y))] \vee \\ &\qquad \exists x[a = see_i(x) \wedge Holding_i(x)] \vee \\ &\qquad a = challenge_i \wedge \text{TRUE}.\end{aligned}$$

In English: sensing actions see_i are explained below, but to sense we assume that i is holding the object. We let $pick_i$ be a physical action, and we require that for object x to be picked up, i is not holding anything else. The other fluent-changing action in the domain is $challenge_i$, such that Σ_{post} has two elements:

$$\Box[a]Holding_i(x) \equiv a = pick_i(x) \vee Holding_i(x).$$

$$\begin{aligned}&\Box[a]Lose_i \equiv Lose_i \vee \\ &\qquad a = challenge_i \wedge (num(card_i) < num(card_j)) \vee \\ &\qquad a = challenge_j \wedge (num(card_j) > num(card_i)).\end{aligned}$$

where $card_i$ is the card that has been dealt to i, num is a rigid function representing the card's number, and $Lose_i$ indicates that i has lost the game. That is, if i challenges, he would win only with the higher number.

Let us now formalize the sensing axioms. When A reads his card, we expect her to discover the number on the card. Actions are public, but despite the fact that B observes A reading her card, B is not expected to discover the contents of A's card. This asymmetry can be captured by letting Σ_{sense} contain the following sentence:

$$\begin{aligned}&\Box[a]SF_i(a) = y \equiv \\ &\qquad \exists x[a = see_i(x) \wedge y = num(x)] \vee \\ &\qquad \neg\exists x[a = see_i(x) \wedge y = \text{NIL}].\end{aligned}$$

In English: i's sensing results for the action $see_i(x)$ informs i about the number on card x, while i's sensing results for every other action returns NIL. That is, NIL is obtained when i senses on physical actions, as well as when i observes j reading a card by means of see_j.

Finally, we stipulate the initial theories, after which we are done. We let Σ_0 be the following:

- $\forall x[\neg Holding_i(x)]$;
- $num(card_A) \neq num(card_B)$;

- $\forall x[num(x) = {}^{\#}1 \lor \ldots \lor num(x) = {}^{\#}52]$;
- ${}^{\#}1 < {}^{\#}2 < {}^{\#}3 < \ldots < {}^{\#}51 < {}^{\#}52$;
- $\neg Lose_A \land \neg Lose_B$.

Here, we are supposing that $card_A$ is the card that A was dealt, while $card_B$ is the one that B was dealt. Basically, Σ_0 says that the numbers on $card_A$ and $card_B$ are different but they are one of $\{1, \ldots, 52\}$ and that initially, no player has lost the game.

Σ_0 represents the initial assumptions of the game. In general, players may have access to additional information. In an unfair setting, for instance, we might imagine that B knows A's card before A does. For our current purposes, however, we will simply assume that Σ_0 is what i believes, as well as what i believes j to believe, at all levels. To now model the real world, let

$$\Upsilon_0 = \Sigma_0 \cup \{num(card_A) = {}^{\#}1, num(card_B) = {}^{\#}52\}.$$

Letting $\Sigma = \Sigma_0 \cup \Sigma_{pre} \cup \Sigma_{post} \cup \Sigma_{sense}$, and letting $\Upsilon = \Sigma \cup \Upsilon_0$, our development leads to a theory of the following form:

$$\Upsilon \land OKnow_\Sigma[A, k] \land OKnow_\Sigma[B, j] \tag{4}$$

Prior to analyzing the entailments of (4), it is convenient to state a lemma regarding how a model of (4) can be constructed. For that, we will use the notion of the *modal depth* of a formula, which refers to the epistemic modalities in the formula.

Definition 13 The *modal depth* of a formula α is defined inductively:

- $modal(\alpha) = 0$ for atomic formulas;
- $modal(\alpha \lor \beta) = max(modal(\alpha), modal(\beta))$;
- $modal(\forall x\alpha) = modal(\alpha)$;
- $modal([t]\alpha) = modal(\alpha)$;
- $modal(\Box\alpha) = modal(\alpha)$;
- $modal(\neg\alpha) = modal(\alpha)$;
- $modal(\boldsymbol{M}_i\alpha) = 1 + modal(\alpha)$ where $\boldsymbol{M}_i \in \{\boldsymbol{K}_i, \boldsymbol{O}_i\}$. ∎

For example, $p \lor [t]q$, where p and q are atoms, is a formula not mentioning epistemic operators and so its modal depth is 0. $\boldsymbol{K}_A\boldsymbol{K}_B p$, in contrast, has a modal depth of 2. Essentially, the modal depth simply counts the epistemic modalities in a formula and completely ignores the indices of these modalities. Not surprisingly, it differs from the i-depth of formulas. For example, the modal depth of $\boldsymbol{K}_A p$ is 1, $|\boldsymbol{K}_A p|_A$ is 1, but $|\boldsymbol{K}_A p|_B$ is 2. In contrast, the modal depth of $\boldsymbol{K}_A\boldsymbol{K}_A p$ is 2, but its A-depth is 1 and its B-depth is 2, as in the case of $\boldsymbol{K}_A p$.

Suppose ϕ is any objective sentence, possibly a basic action theory. Let us denote the set of worlds $\{w \mid w \models \phi\}$ as $\mathcal{W}_\phi$. Further, let $e^1_\phi = \mathcal{W}_\phi \times \{\{\}\}$. Let $e^k_\phi = \{(w, e^{k-1}_\phi) \mid w \in \mathcal{W}_\phi\}$ be defined inductively. Then,

Lemma 14 *Suppose ϕ is an objective sentence. Suppose w is any world and $e_{\phi}{}_A^k$ and $e_{\phi}{}_B^j$ are constructed as above. Then $e_{\phi}{}_A^k, e_{\phi}{}_B^j, w \models OKnow_\phi[A,k] \wedge OKnow_\phi[B,j]$.*

Proof: The proof is by an induction on the *modal depth* of the background theory, which Definition 13 provides. First note that when the modal depth of the background theory is l, then we have a sentence of the form $OKnow_\phi[A,k] \wedge OKnow_\phi[B,j]$ such that $k \leq l$, $j \leq l$ and k or j is l.

Since $OKnow_\phi[i,k]$ is interpreted wrt i's epistemic state, we can treat the A-subjective and B-subjective formulas of the background theory individually. The base case is for theories of modal depth 1, where we are considering a sentence of the form $\boldsymbol{O}_i\phi$. To prove the base case, consider any world w'. Clearly $w' \simeq_{\langle\rangle} w$ by definition. By construction, $(w', \{\}) \in e_{\phi}{}_A^1$ iff $w' \models \phi$. Therefore $e_{\phi}{}_A^1, \{\}, w \models \boldsymbol{O}_A\phi$. Analogously for $e_{\phi}{}_B^1$.

Suppose that the lemma holds for background theories of modal depth $k-1$, that is, $e_{\phi}{}_A^{k-1}$ satisfies $OKnow_\phi[A,k-1]$. This is analogously stated for B. Let $(w', e_{\phi}{}_B^{k-1})$ be any k-structure in $e_{\phi}{}_A^k$. By construction $w' \models \phi$. By induction hypothesis, $\{\}, e_{\phi}{}_B^{k-1}, w' \models OKnow_\phi[B,k-1]$. That is, by construction, $(w', e_{\phi}{}_B^{k-1}) \in e_{\phi}{}_A^k$ iff $\{\}, e_{\phi}{}_B^{k-1}, w' \models \phi \wedge OKnow_\phi[B,k-1]$. Therefore $e_{\phi}{}_A^k, \{\}, w \models \boldsymbol{O}_A(\phi \wedge OKnow_\phi[B,k-1])$, that is, $e_{\phi}{}_A^k, \{\}, w \models OKnow_\phi[A,k]$. ∎

Using this lemma, we now consider some properties of (4):

Proposition 15 *The following sentences are entailed by the sentence* (4)*, with $k > 1$ and $j > 1$.*

1. $\neg\boldsymbol{K}_A(num(card_A) = {}^{\#}1)$.

   Initially, A does not know the details of her card. (That is, non-beliefs are obtained via only knowing.)

2. $[pick_A(card_A)][see_A(card_A)]\boldsymbol{K}_A(num(card_A) = {}^{\#}1)$.

   After sensing, A knows he has the lowest number.

3. $[pick_A(card_A)][see_A(card_A)]\boldsymbol{K}_B\exists x\boldsymbol{K}_A(num(card_A) = x)$.

   After B observes A reading his card, B knows that A knows what $card_A$ holds for him. That is, B has de dicto knowledge about A's knowledge.

4. $[pick_A(card_A)][see_A(card_A)]\neg\exists x\boldsymbol{K}_B(num(card_A) = x)$.

   But it is not the case that B knows A's card when he observes A sensing. That is, B does not have de re knowledge about the card.

5. $[pick_A(card_A)][see_A(card_A)]\boldsymbol{K}_A\neg\exists x\boldsymbol{K}_B(num(card_A) = x)$.

   Moreover, A knows that B does not know her card.

6. $[pick_A(card_A)][see_A(card_A)][pick_B(card_B)][see_B(card_B)]\boldsymbol{K}_i([challenge_B]Lose_A)$ *where $i \in \{A, B\}$.*

   After sensing, both A and B believe that A would lose the game if challenged by B.

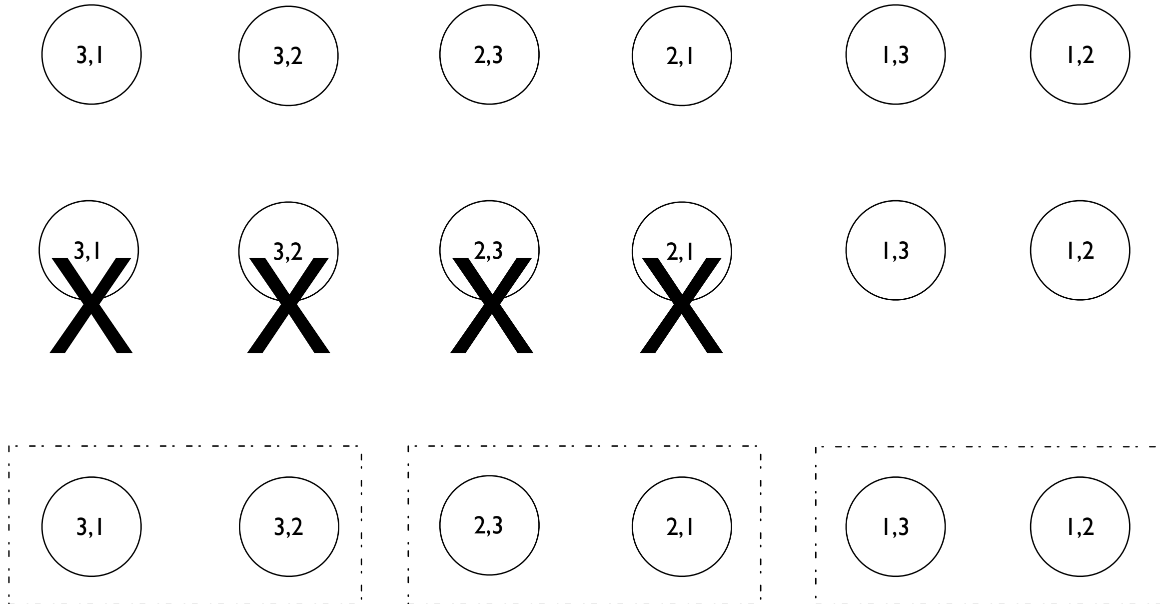

Figure 2: This depicts the compatibility of worlds after actions, shown wrt a 3-card deck for simplicity's sake. Here, worlds are characterized in terms of the numbers on the cards, and therefore, they are simply labeled (n, m), where n denotes the number on A's card and m denotes the number on B's card at the world. The first line represent's A's uncertainty initially, and the second after A senses $^{\#}1$, in which case all worlds where she does not have that card are discarded. The third represents B's belief about A's epistemic state after he observes A sensing her card. That is, while he does not know which card A has, he does know that A considers only two worlds possible, grouped as shown, without knowing which group represents truth.

Proof: Let $M = (e_A^k, e_B^j, w)$ be *any* model of (4). It is easy to see that e_A^k and e_B^j would be as in Lemma 14, and w is any world satisfying Υ from (4). Below, we let r denote $pick_A(card_A)$ and let r' denote $see_A(card_A)$.

1. Assume the contrary. Suppose that $e_A^k, e_B^j, w \models \boldsymbol{K}_A(num(card_A) = {}^{\#}1)$. Then for all $(w', e_B^{k-1}) \in e_A^k$, we get $e_A^k, e_B^{k-1}, w' \models (num(card_A) = {}^{\#}1)$.

   Now, observe that Σ_0 only says that the value of $num(card_A) \in \{1, \ldots, 52\}$. Thus, by construction (and the definition of $\mathcal{W}$), there are worlds $w^* \in \mathcal{W}_\Sigma$ where (say) $w^* \models (num(card_A) = {}^{\#}2)$, $w^* \simeq^A_{\langle\rangle} w$, and $(w^*, e_B^{k-1}) \in e_A^k$ for some e_B^{k-1}. This is a contradiction.

2. After A executes $r \cdot r'$, it follows that only those worlds $w' \in \mathcal{W}_\Sigma$ such that $w'[SF_A(r'), r] = w[SF_A(r'), r] = {}^{\#}1$ are considered when evaluating A-subjective formulas. (These are worlds that agree on the number on A's card with the real world.) Therefore $e_A^k, e_B^j, w, r \cdot r' \models \boldsymbol{K}_A(num(card_A) = {}^{\#}1)$ since for every $(w', e_B^{k-1}) \in e_A^k$ such that $w' \simeq^A_{r \cdot r'} w$, $e_A^k, e_B^{k-1}, w', r \cdot r' \models (num(card_A) = {}^{\#}1)$ by the definition of the semantics and the sensing axioms Σ_{sense}.

3. Consider any $(w^*, e_A^{j-1}) \in e_B^j$. We get that $w^* \simeq^B_{r \cdot r'} w$ iff $w^*[SF_B(r'), r] = w[SF_B(r'), r]$. Since $SF_B(r) = SF_B(r') = \text{NIL}$ in all worlds satisfying Σ, for item 3 to hold it must follow that for every $(w', e_A^{j-1}) \in e_B^j, e_A^{j-1}, e_B^j, w', r \cdot r' \models \exists x \boldsymbol{K}_A(num(card_A) = x)$, and so $e_A^{j-1}, e_B^j, w', r \cdot r' \models \boldsymbol{K}_A(num(card_A) = n)$ for some n. Using arguments from item 2, this is easily shown to be the case. That is, $e_A^{j-1}, e_B^j, w' \models \boldsymbol{K}_A(num(card_A) = n)$ iff for every $w'' \simeq^A_{r \cdot r'} w'$, if $(w'', e_B^{j-2}) \in e_A^{j-1}$, then $e_A^{j-1}, e_B^{j-2}, w'' \models (num(card_A) = n)$. This holds because $w'' \simeq^A_{r \cdot r'} w'$ will hold when $w'', r \cdot r' \models (num(card_A) = n)$.

   The intuitive argument is as follows. Suppose B only considered j-structures (w, e_A^{j-1}) possible, where w is the real world. Then he would be able to infer $card_A$'s number. But since his epistemic state is $\{(w', e_A^{j-1}), (w'', e_A^{j-1}), \ldots\}$ he believes at each of the worlds w' that A knows his number as well as what this is, but he does not know of which of these is the real world. In effect, there are some structures that inform B that A's card is $^{\#}1$, and there are others that inform him that A's card is a different number, leaving him uncertain. For the case of a 3-card deck, Figure 2 illustrates this development.

4. This follows from the arguments for the previous item. Basically, for every $w' \in \mathcal{W}_\Sigma$, $w'[SF_B(r), r] = \text{NIL} = w[SF_B(r), r]$. When evaluating B-subjective formulas every $(w', e_A^{j-1}) \in e_B^j$ is considered, including ones where (say) $w', r \cdot r' \models (num(card_A) = {}^{\#}2)$. Thus $e_A^k, e_B^j, w, r \not\models \boldsymbol{K}_B(num(card_A) = {}^{\#}1)$.

5. Consider any $(w', e_B^{k-1}) \in e_A^k$ such that $w' \simeq^A_{r \cdot r'} w$. By the arguments from item 4, it follows that
$$e_A^k, e_B^{k-1}, w', r \cdot r' \not\models \exists x \boldsymbol{K}_B(num(card_A) = x).$$
   Therefore, $e_A^k, e_B^j, w, r \cdot r' \models \boldsymbol{K}_A \neg \exists x \boldsymbol{K}_B(num(card_A) = x)$.

6. This property follows from logical deduction. After sensing, both players know their own cards. A has the lowest number, while B has the highest. Since both agents know that their numbers are unique, and these numbers are one of $\{1, \ldots, 52\}$, both infer that A would lose after being challenged. ∎

4. Regression

A fundamental reasoning task in dynamic domains is *projection* (Reiter, 2001), where we are to infer whether α holds after a sequence of actions $a_1, \ldots, a_k$ is executed:

$$\Upsilon \models [a_1] \ldots [a_k]\alpha.$$

Reiter (2001) developed an important solution to the projection problem in the situation calculus called *regression*. The idea is to reduce a query α about the future to a query α' about the initial situation by successively replacing fluents in α by the *rhs* of the successor state axioms until the resulting sentence α' contains no more actions. We then need to only verify whether α' is entailed by the initial theory.

In the context of multiagent systems, we might, for example, be interested in reasoning about entailments about a background theory such as (3) that stipulates the beliefs of agents in the application domain:

$$(3) \models [a_1] \ldots [a_k]\phi$$

where ϕ perhaps mentions belief operators.

Reiter's (2001) results were extended by Scherl and Levesque (2003) to handle knowledge in the situation calculus, which was further shown to carry over to $\mathcal{ES}$ (Lakemeyer & Levesque, 2011). In this section, we generalize these results for background theories of the form (3) involving multiagent only knowing operators. We first recap the regression of objective formulas from $\mathcal{ES}$.

4.1 Regressing Objective Formulas

Without any loss of generality, we assume that the query α is syntactically reformulated as follows:

1. quantifiers use distinct variables, and we say such formulas are *rectified*;
2. formulas are in a certain normal form called NF (defined below).

After applying these transformations, the query becomes amenable to regression. The first syntactic manipulation is required because of the way regression handles quantifiers, which can lead to incorrect transformations if the variables are not distinct. The second is required for giving a simple formulation of regression.

Definition 16 A formula α is in NF if every function symbol f in α occurs only in equality expressions of the form $f(t_1, \ldots, t_k) = t'$, where t_i and t' are either variables or names. ∎

It is immediate to verify that every formula can be rewritten to one in NF, and this transformation is linear in the size of the formula. For instance, $f(g(x)) = f'(x)$ is equivalent to $\exists y, u.\ f(y) = u \wedge f'(x) = u \wedge g(x) = y$. Further, by this definition, if a term t appears either as an argument for a function or as an action operator $[t]$, then it follows that it is either an (action) name or a variable. In the following we will use σ to denote sequences that consist of action variables or action names.

Lakemeyer and Levesque (2011) define the regression operator $\mathcal{R}$, which is applicable to any bounded objective formula.[12] If such a formula is not rectified or not in NF, it is transformed to a formula satisfying these conditions.

Definition 17 Define $\mathcal{R}[\alpha]$, the regression of a bounded basic formula α wrt Υ, to be the fluent formula $\mathcal{R}[\langle\rangle, \alpha]$. For any sequence of action names or variables σ, $\mathcal{R}[\sigma, \alpha]$ is defined inductively:

1. $\mathcal{R}[\sigma, t_1 = t_2] = (t_1 = t_2)$ if t_1 and t_2 do not mention functional fluents;
2. $\mathcal{R}[\sigma, \forall x\alpha] = \forall x\mathcal{R}[\sigma, \alpha]$;
3. $\mathcal{R}[\sigma, \alpha \vee \beta] = \mathcal{R}[\sigma, \alpha] \vee \mathcal{R}[\sigma, \beta]$;
4. $\mathcal{R}[\sigma, \neg\alpha] = \neg\mathcal{R}[\sigma, \alpha]$;
5. $\mathcal{R}[\sigma, [t]\alpha] = \mathcal{R}[\sigma \cdot t, \alpha]$;
6. $\mathcal{R}[\sigma, Poss(t)] = \mathcal{R}[\sigma, \pi_t^a]$;

12. Roughly speaking, these correspond to the class of formulas that Reiter (2001) deems *regressable* in the situation calculus.

7. $\mathcal{R}[\sigma, G(t_1, \ldots, t_k)] = G(t_1, \ldots, G_k)$ for rigid predicate G;

8. $\mathcal{R}[\sigma, F(t_1, \ldots, t_k)]$ for fluent predicate F is defined inductively on σ:

   (a) $\mathcal{R}[\langle\rangle, F(t_1, \ldots, t_k)] = F(t_1, \ldots, t_k)$;

   (b) $\mathcal{R}[\sigma \cdot t, F(t_1, \ldots, t_k)] = \mathcal{R}[\sigma, {\gamma_F}^{a\ x_1\ \ldots\ x_k}_{t\ t_1\ \ldots\ t_k}]$;

9. $\mathcal{R}[\sigma, f(t_1, \ldots, t_k) = t']$ for fluent function f is defined inductively by:

   (a) $\mathcal{R}[\langle\rangle, f(t_1, \ldots, t_k) = t'] = (f(t_1, \ldots, t_k) = t')$;

   (b) $\mathcal{R}[\sigma \cdot t, f(t_1, \ldots, t_k) = t'] = \mathcal{R}[\sigma, \exists y.\ (\gamma_f)^{a x_1\ \ldots\ x_k}_{t\ t_1\ \ldots\ t_k} \wedge y = t']$. ∎

Note that this definition includes π, γ_F and γ_f which are the *rhs* of the precondition axiom and the successor state axioms from Υ.

The main result regarding Definition 17 is that the evaluation of objective bounded sentences reduces to a query about the initial theory.

Theorem 18 *(Lakemeyer & Levesque, 2011) Let Υ be a basic action theory, whose initial theory is Υ_0, and let α be any objective bounded sentence. Then $\mathcal{R}[\alpha]$ is a fluent sentence and satisfies:*

$$\Upsilon \models \alpha \quad \textit{iff} \quad \Upsilon_0 \models \mathcal{R}[\alpha].$$

4.2 Regressing Multiagent Beliefs

Let us now consider the more general case of regression for bounded sentences mentioning belief operators. This first needs the equivalent of a successor state axiom for knowledge, which will tell us how knowledge can be regressed wrt an action. The following theorem generalizes a similar result by Lakemeyer and Levesque (2004) to the many agent case.

Theorem 19 (Successor State Axiom for Knowledge.)

$$\begin{aligned}&\models \Box[a]\boldsymbol{K}_i(\alpha) \equiv \\ &\qquad \exists x.\ SF_i(a) = x \wedge \boldsymbol{K}_i(SF_i(a) = x \supset [a]\alpha).\end{aligned}$$

Proof: Let i be A, with the other case being symmetric. For the only-if direction, suppose that $e^k_A, e^j_B, w, z \models [r]\boldsymbol{K}_A\alpha^a_r$ for an action name r. Abbreviate α^a_r as α'. Suppose that $e^k_A, e^j_B, w, z \models SF_A(r) = n$. It then suffices to show that $e^k_A, e^j_B, w, z \models \boldsymbol{K}_A(SF_A(r) = n \supset [r]\alpha')$.

So suppose $(w', e^{k-1}_B) \in e^k_A$ and $w'[SF_A(r), z] = n$. Since $w' \simeq^A_{z\cdot r} w$, it follows by assumption that $e^k_A, e^{k-1}_B, w', z \cdot r \models \alpha'$, *i.e.* $e^k_A, e^{k-1}_B, w', z \models [r]\alpha'$. Thus $e^k_A, e^{k-1}_B, w', z \models SF_A(r) = n \supset [r]\alpha'$, and it follows then that $e^k_A, e^j_B, w', z \models \boldsymbol{K}_A(SF_A(r) = n \supset [r]\alpha')$.

Conversely, suppose that $e^k_A, e^j_B, w, z \models SF_A(r) = n \wedge \boldsymbol{K}_A(SF_A(r) = n \supset [r]\alpha')$. We now need to show that $e^k_A, e^j_B, w, z \models [r]\boldsymbol{K}_A(\alpha')$, *i.e.* for all $(w', e^{k-1}_B) \in e^k_A$ such that $w' \simeq^A_z w$, $e^k_A, e^{k-1}_B, w', z{\cdot}r \models \alpha'$.

Suppose $w' \simeq^A_{z\cdot r} w$, *i.e.* $w'[SF_A(r), z] = n$ and $(w', e^{k-1}_B) \in e^k_A$ for some e^{k-1}_B. Then by assumption, $e^k_A, e^{k-1}_B, w', z \cdot r \models \alpha'$. So $e^k_A, e^{k-1}_B, w', z \models [r]\alpha$, from which we get $e^k_A, e^j_B, w, z \models \boldsymbol{K}_A([r]\alpha')$. ∎

This theorem essentially says that knowledge after an action depends on what was known before, and what the future would look like contingent on the sensing result. Note that this theorem is not a stipulation of the action theory (Scherl & Levesque, 2003), but a theorem of the logic.

We mentioned earlier that in the case of (3), we need three basic action theories Υ, Σ and Σ'. The idea behind regression is to transform objective formulas wrt Υ, while subjective ones are regressed wrt Σ and Σ'. Consequently, $\mathcal{R}$ is defined wrt Υ, Σ and Σ'. More precisely, we define a regression operator $\mathcal{R}[\Upsilon, \Sigma, \Sigma', \sigma, \alpha]$ wrt a basic action theory Υ for what is true in the real world, a basic action theory Σ for what *A* believes at all levels, and a basic action theory Σ' for what *B* believes at all levels, as expected by (3).

Definition 20 We define $\mathcal{R}[\Upsilon, \Sigma, \Sigma', \alpha]$, the *regression* of a bounded basic formula α wrt Υ, Σ and Σ', to be $\mathcal{R}[\Upsilon, \Sigma, \Sigma', \langle\rangle, \alpha]$. For a given sequence of action names or variables σ, we define $\mathcal{R}[\Upsilon, \Sigma, \Sigma', \sigma, \alpha]$ inductively by:

1.-9. See Definition 17. (Note that this definition uses the *rhs* of the precondition axiom and the successor state axioms from Υ.)

10. $\mathcal{R}[\Upsilon, \Sigma, \Sigma', \sigma, SF_A(t) = t'] = \mathcal{R}[\Upsilon, \Sigma, \Sigma', \sigma, \varphi_A{}^{a}_{t}{}^{x}_{t'}]$ which uses the *rhs* of the sensing axioms from Σ;

11. $\mathcal{R}[\Upsilon, \Sigma, \Sigma', \sigma, SF_B(t) = t'] = \mathcal{R}[\Upsilon, \Sigma, \Sigma', \sigma, \varphi_B{}^{a}_{t}{}^{x}_{t'}]$ which uses the *rhs* of the sensing axioms from Σ';

12. $\mathcal{R}[\Upsilon, \Sigma, \Sigma', \sigma, \boldsymbol{K}_A\alpha]$ is defined inductively on σ by:
    (a) $\mathcal{R}[\Upsilon, \Sigma, \Sigma', \langle\rangle, \boldsymbol{K}_A\alpha] = \boldsymbol{K}_A(\mathcal{R}[\Sigma, \Sigma, \Sigma, \langle\rangle, \alpha])$;
    (b) $\mathcal{R}[\Upsilon, \Sigma, \Sigma', \sigma \cdot t, \boldsymbol{K}_A\alpha] = \mathcal{R}[\Upsilon, \Sigma, \Sigma', \sigma, \beta^a_t]$, where β is *rhs* of the equivalence in Theorem 19 for the agent index *A*.

13. $\mathcal{R}[\Upsilon, \Sigma, \Sigma', \sigma, \boldsymbol{K}_B\alpha]$ is defined inductively on σ by:
    (a) $\mathcal{R}[\Upsilon, \Sigma, \Sigma', \langle\rangle, \boldsymbol{K}_B\alpha] = \boldsymbol{K}_B(\mathcal{R}[\Sigma', \Sigma', \Sigma', \langle\rangle, \alpha])$;
    (b) $\mathcal{R}[\Upsilon, \Sigma, \Sigma', \sigma \cdot t, \boldsymbol{K}_B\alpha] = \mathcal{R}[\Upsilon, \Sigma, \Sigma', \sigma, \beta^a_t]$, where β is *rhs* of the equivalence in Theorem 19 for the agent index *B*. ∎

The regression operator in the multiagent case works as follows. At the initial situation, regressing $\boldsymbol{K}_A\alpha$ is equivalent to regressing α wrt the basic action theory Σ that *A* believes at all levels. Similarly, at the initial situation, regressing $\boldsymbol{K}_B\alpha$ is equivalent to regressing α wrt the basic action theory Σ' that *B* believes at all levels. More generally, if we are regressing $\boldsymbol{K}_i\alpha$ wrt an action sequence $\sigma \cdot t$, then this is equivalent to regressing the *rhs* of Theorem 19 wrt σ by first substituting *t*.

For simplicity, we often write $\mathcal{R}[\sigma, \alpha]$ instead of $\mathcal{R}[\Upsilon, \Sigma, \Sigma', \sigma, \alpha]$. We are now ready to prove the main result of this section for bounded basic sentences:[13]

13. Roughly speaking, these correspond to the class of *regressable formulas* in the epistemic situation calculus (Scherl & Levesque, 2003).

Theorem 21 *Suppose α is a bounded basic sentence of maximal A, B-depth k, j. Let Υ, Σ and Σ' be basic action theories. Then $\mathcal{R}[\langle\rangle, \alpha]$ is a static sentence and satisfies:*

$$\Upsilon \wedge \psi \models \alpha \quad \textit{iff} \quad \Upsilon_0 \wedge \psi_0 \models \mathcal{R}[\langle\rangle, \alpha]$$

where $\psi = \mathit{OKnow}_{\Sigma}[A, k] \wedge \mathit{OKnow}_{\Sigma'}[B, j]$
$\psi_0 = \mathit{OKnow}_{\Sigma_0}[A, k] \wedge \mathit{OKnow}_{\Sigma_0'}[B, j]$.

That is, we solve projection which is the task of verifying whether α is entailed by regressing α and verifying that is an entailment of the conjunction of what is true initially and each agent only knowing their initial beliefs. The proof for this theorem is provided in Appendix A.

Readers will have noticed that the theorem assumes a background theory where A has beliefs to level k and B has beliefs to level j, given a query whose maximal A, B-depth is k, j. This syntactic restriction is essential for our relatively simple regression operator to be well-defined. To see that, suppose we are interested in verifying whether $\boldsymbol{K}_A \boldsymbol{K}_B[r]\alpha$ is entailed by $\boldsymbol{O}_A(\Sigma)$, where Σ is a basic action theory. By the definition of the regression operator given above, evaluating the query reduces to regressing $[r]\alpha$ wrt Σ, but this is not a correct transformation because A does not have any beliefs about B's knowledge of the world. In fact, the formula $\boldsymbol{K}_A \boldsymbol{K}_B[r]\alpha$ does not seem amenable to regression wrt $\boldsymbol{O}_A(\Sigma)$ since it is simply not clear how one should regress the subformula $\boldsymbol{K}_B[r]\alpha$. But now note that the formula $\boldsymbol{K}_A \boldsymbol{K}_B[r]\alpha$ is of depth 2 and that the transformation is indeed correct wrt initial knowledge for A of at least depth 2, such as $\boldsymbol{O}_A(\Sigma \wedge \boldsymbol{O}_B\Sigma)$.

Readers will also notice that we are restricting the regression operator to bounded basic sentences. There are at least two reasons for this limitation. First, note that the language is not expressive enough to refer to only knowing in non-initial situations; if an agent only knows a basic action theory, one presumes that after an action the agent only knows another basic action theory. Regressing the latter should intuitively lead to a sentence that talks about what was only known before the action was executed, and this currently cannot be expressed in the language. Second, note that a basic action theory contains sentences such as the successor state axioms which are not bounded. So, if after an action we are left with a formula of the form $\boldsymbol{O}_i(\alpha)$, where α by the above argument would contain sentences that are not bounded, then this α would not be regressable. This is because Theorem 18 is limited to regressing bounded formulas. Nevertheless, the regression operator covers the same class of formulas as considered by Scherl and Levesque (2003), and is sufficient for most practical purposes.

Example 22 We illustrate regression using the card game. Suppose we are interested in checking whether (4) from Section 3 entails the following sentence:

$$[\mathit{pick}_A(\mathit{card}_A)][\mathit{see}_A(\mathit{card}_A)](\boldsymbol{K}_A(\mathit{num}(\mathit{card}_A) = {}^{\#}1) \wedge \neg\boldsymbol{K}_B(\mathit{num}(\mathit{card}_A) = {}^{\#}1)). \tag{5}$$

That is, after A picks her card and senses, she knows of her own card while B does not learn this. Use r for $\mathit{pick}_A(\mathit{card}_A)$, r' for $\mathit{see}_A(\mathit{card}_A)$ and α for $(\mathit{num}(\mathit{card}_A) = {}^{\#}1)$. Begin with

$$\begin{aligned}
&\mathcal{R}[\Upsilon, \Sigma, \Sigma, r \cdot r', \boldsymbol{K}_A\alpha \wedge \neg\boldsymbol{K}_B\alpha] \\
&\quad = \mathcal{R}[\Upsilon, \Sigma, \Sigma, r \cdot r', \boldsymbol{K}_A\alpha] \wedge \mathcal{R}[\Upsilon, \Sigma, \Sigma, r \cdot r', \neg\boldsymbol{K}_B\alpha]
\end{aligned}$$

Consider $[r][r']\boldsymbol{K}_A\alpha$. We have:

$$\begin{aligned}
&\mathcal{R}[\Upsilon, \Sigma, \Sigma, r\cdot r', \boldsymbol{K}_A\alpha]\\
&\quad = \mathcal{R}[\Upsilon, \Sigma, \Sigma, r, \exists x(SF_A(r') = x \wedge \boldsymbol{K}_A(SF_A(r') = x \supset [r']\alpha))]\\
&\quad = \exists x.\ \mathcal{R}[\Upsilon, \Sigma, \Sigma, r, SF_A(r') = x] \wedge \mathcal{R}[\Upsilon, \Sigma, \Sigma, r, \boldsymbol{K}_A(SF_A(r') = x \supset [r']\alpha)]\\
&\quad = \exists x.\ \mathcal{R}[\Upsilon, \Sigma, \Sigma, r, num(card_A) = x] \wedge\\
&\qquad\qquad \mathcal{R}[\Upsilon, \Sigma, \Sigma, \langle\rangle, \exists y.\ SF_A(r) = y \wedge \boldsymbol{K}_A(SF_A(r) = y \supset [r]\beta)]\\
&\quad = \exists x.\ num(card_A) = x \wedge \exists y.\ y = \mathrm{NIL}\ \wedge \boldsymbol{K}_A(\mathcal{R}[\Sigma, \Sigma, \Sigma, \langle\rangle, SF_A(r) = y \supset [r]\beta])\\
&\quad = \exists x.\ num(card_A) = x \wedge \exists y.\ y = \mathrm{NIL} \wedge\\
&\qquad\qquad \boldsymbol{K}_A(y = \mathrm{NIL} \supset (num(card_A) = x \supset num(card_A) = {}^{\#}1))
\end{aligned}$$

Here, β denotes $(SF_A(r') = x \supset [r']\alpha)$.

These reductions mostly involve repeated applications of the knowledge successor state axiom, over logical connectives. Then, in step 4, when regressing knowledge wrt $\langle\rangle$, we replace all basic action theories in the $\mathcal{R}$ operator with the one that A believes, as expected by Rule 12(a) of Definition 20. Regarding $\mathcal{R}$'s result, since Υ_0 contains $num(card_A) = {}^{\#}1$, it is not hard to see that

$$\Upsilon_0 \wedge OKnow_{\Sigma_0}[A, k] \wedge OKnow_{\Sigma_0}[B, j] \models \mathcal{R}[\Upsilon, \Sigma, \Sigma, r\cdot r', \boldsymbol{K}_A\alpha].$$

Simplifying $\mathcal{R}[\Upsilon, \Sigma, \Sigma, r\cdot r', \neg\boldsymbol{K}_B\alpha]$ is analogous, the only dissimilarity arising from regressing B's beliefs wrt r', which obtains the sensing result NIL:

$$\begin{aligned}
&\mathcal{R}[\Upsilon, \Sigma, \Sigma, r\cdot r', \neg\boldsymbol{K}_B\alpha]\\
&\quad = \neg\mathcal{R}[\Upsilon, \Sigma, \Sigma, r\cdot r', \boldsymbol{K}_B\alpha]\\
&\quad = \neg[\exists x.\ x = \mathrm{NIL} \wedge \exists y.\ y = \mathrm{NIL} \wedge \boldsymbol{K}_B(y = \mathrm{NIL} \supset (x = \mathrm{NIL} \supset num(card_A) = {}^{\#}1))].
\end{aligned}$$

One may verify that

$$\Upsilon_0 \wedge OKnow_{\Sigma_0}[A, k] \wedge OKnow_{\Sigma_0}[B, j] \models \neg\mathcal{R}[\Upsilon, \Sigma, \Sigma, r\cdot r', \boldsymbol{K}_B\alpha].$$

Therefore,

$$\Upsilon_0 \wedge OKnow_{\Sigma_0}[A, k] \wedge OKnow_{\Sigma_0}[B, j] \models \mathcal{R}[\Upsilon, \Sigma, \Sigma, r\cdot r', \boldsymbol{K}_A\alpha \wedge \neg\boldsymbol{K}_B\alpha].$$

By means of Theorem 21, this allows us to conclude that the query (5) is indeed entailed by (4). ∎

Example 23 We will regress nested beliefs in this example. Suppose we are interested in checking whether (4) from Section 3 entails the following sentence:

$$[pick_A(card_A)][see_A(card_A)]\boldsymbol{K}_A\neg\boldsymbol{K}_B(num(card_A) = {}^{\#}1). \tag{6}$$

As we have done so above, let r denote $pick_A(card_A)$, r' denote $see_A(card_A)$, and α denote $num(card_A) = {}^{\#}1$. Then:

$$
\begin{aligned}
&\mathcal{R}[\Upsilon, \Sigma, \Sigma, r \cdot r', \boldsymbol{K}_A \neg \boldsymbol{K}_B \alpha] \\
&\quad = \mathcal{R}[\Upsilon, \Sigma, \Sigma, r, \exists x (SF_A(r') = x \wedge \boldsymbol{K}_A(SF_A(r') = x \supset [r'] \neg \boldsymbol{K}_B \alpha))] \\
&\quad = \exists x.\ \mathcal{R}[\Upsilon, \Sigma, \Sigma, r, num(card_A) = x] \wedge \mathcal{R}[\Upsilon, \Sigma, \Sigma, r, \boldsymbol{K}_A(x = num(card_A) \supset [r'] \neg \boldsymbol{K}_B \alpha)] \\
&\quad = \exists x.\ num(card_A) = x \wedge \\
&\qquad\qquad \mathcal{R}[\Upsilon, \Sigma, \Sigma, \langle\rangle, \exists y (SF_A(r) = y \wedge \boldsymbol{K}_A(SF_A(r) = y \supset [r]\beta))] \\
&\quad = \exists x.\ num(card_A) = x \wedge \exists y.\ y = \mathrm{NIL} \wedge \boldsymbol{K}_A(\mathcal{R}[\Sigma, \Sigma, \Sigma, \langle\rangle, y = \mathrm{NIL} \supset [r]\beta]).
\end{aligned}
$$

Here, β denotes $(num(card_A) = x \supset [r'] \neg \boldsymbol{K}_B \alpha)$.

The above reduction leads to:

$$\exists x.\ num(card_A) = x \wedge \exists y.\ y = \mathrm{NIL} \wedge \boldsymbol{K}_A(y = \mathrm{NIL} \supset \mathcal{R}[\Sigma, \Sigma, \Sigma, \langle\rangle, [r]\beta]). \tag{7}$$

Let us consider $\mathcal{R}[\Sigma, \Sigma, \Sigma, \langle\rangle, [r]\beta]$. We have (on simplification):

$$
\begin{aligned}
&\mathcal{R}[\Sigma, \Sigma, \Sigma, \langle\rangle, [r]\beta] \\
&\quad = (num(card_A) = x) \supset \mathcal{R}[\Sigma, \Sigma, \Sigma, \langle\rangle, [r][r'] \neg \boldsymbol{K}_B \alpha].
\end{aligned}
$$

Following the reduction of $\mathcal{R}[\Sigma, \Sigma, \Sigma, \langle\rangle, [r][r'] \neg \boldsymbol{K}_B \alpha]$ as done in the previous example, it can be shown that

$$\Upsilon_0 \wedge OKnow_{\Sigma_0}[A, k] \wedge OKnow_{\Sigma_0}[B, j] \models (7).$$

Using the regression property, that is, Theorem 21, we conclude that (6) is entailed by (4). Therefore, we are done. (We reiterate that A has knowledge about B's non-beliefs because A has beliefs about what B only knows.) ∎

Analogously, other entailments from Proposition 15 can be verified using regression.

While regression allows us to reduce questions about knowledge and action to queries about initial beliefs, in the next section we go even further and replace reasoning about knowledge by classical first-order reasoning.

4.3 A Representation Theorem

The representation theorem is a result by means of which reasoning about knowledge is reduced to first-order theorem proving. The presentation below generalizes the single agent variant (Lakemeyer & Levesque, 2004).

The basic idea is to substitute believed sentences with their instances. For example, suppose all that i believes is the sentence ϕ:

$$\phi = \{Smaller(n_2, n_1), Smaller(n_3, n_1) \vee Smaller(n_4, n_1)\}.$$

That is, in a blocks world domain: n_2 is smaller than n_1, and n_3 is smaller than n_1 or n_4 is smaller than n_1. Supposing we ask:

$$\boldsymbol{K}_i \exists x.\ (Smaller(x, n_1) \wedge \neg \boldsymbol{K}_i Smaller(x, n_1)) \tag{8}$$

That is, does ϕ know of a block that is smaller than n_1, but does not know which one? The answer is certainly *yes* because the list of smaller blocks known is incomplete, except for n_2. The essential step is to replace $\boldsymbol{K}_i Smaller(x, n_1)$ with $x = n_2$. Then, it can be shown that the query reduces to verifying if $\exists x. (Smaller(x, n_1) \wedge x \neq n_2)$ is entailed by ϕ.

To make this intuition precise, we first define a procedure $\textsc{Res}[\alpha, \phi]$, introduced originally by Levesque (1990), to obtain the known instances of α that are entailed by ϕ, where both α and ϕ are fluent formulas. When α does not mention free variables, then $\textsc{Res}$ checks whether ϕ entails the sentence α.

Definition 24 Let α be a fluent formula, and ϕ is a fluent sentence. Let $n_1, \ldots, n_k$ be all the names occurring in ϕ and α and n' is a name not occurring in ϕ or α. Then, $\textsc{Res}[\alpha, \phi]$ is defined as:

1. If α has no free variables, then $\textsc{Res}[\alpha, \phi]$ is TRUE if $\phi \models \alpha$ and FALSE otherwise.
2. If x is a free variable in α, then $\textsc{Res}[\alpha, \phi]$ is defined as:

$$\begin{aligned}
&((x = n_1) \wedge \textsc{Res}[\alpha^x_{n_1}, \phi]) \vee \\
&\qquad\qquad\qquad\qquad \ldots \vee \\
&((x = n_k) \wedge \textsc{Res}[\alpha^x_{n_k}, \phi]) \vee \\
&((x \neq n_1) \wedge \ldots \wedge (x \neq n_k) \wedge \textsc{Res}[\alpha^x_{n'}, \phi]^{n'}_x). \blacksquare
\end{aligned}$$

For instance, if α was $Smaller(x, n_1)$ and ϕ is as above, then $\textsc{Res}[\alpha, \phi]$ would simplify to:

$$(x = n_2) \wedge \textsc{Res}[\alpha^x_{n_2}, \phi]$$

where, further, $\textsc{Res}[\alpha^x_{n_2}, \phi]$ is TRUE because $\phi \models Smaller(n_2, n_1)$.

Given any formula $\boldsymbol{K}_i\alpha$ and a ϕ that is only known by i, the idea is to reason about knowledge by utilizing $\textsc{Res}$. Of course, as discussed earlier, in the multiagent case we have to account for knowledge bases at the different levels, that is, by addressing background theories of the form (3). So we proceed as follows. Let ϕ and ϕ' denote the initial theories (fluent sentences) believed by A and B at all levels respectively. Then, given any bounded basic sentence, we first use regression to obtain a static basic sentence. For any static basic α, define $||\alpha||_{\phi,\phi'}$ as follows:

Definition 25 Let ϕ and ϕ' be fluent sentences, and α and β be static basic sentences. Then we define the fluent sentence $||\alpha||_{\phi,\phi'}$ by:

1. $||\alpha||_{\phi,\phi'} = \alpha$ if α is objective;
2. $||\neg\alpha||_{\phi,\phi'} = \neg||\alpha||_{\phi,\phi'}$;
3. $||\alpha \vee \beta||_{\phi,\phi'} = ||\alpha||_{\phi,\phi'} \vee ||\beta||_{\phi,\phi'}$;
4. $||\forall x\alpha||_{\phi,\phi'} = \forall x||\alpha||_{\phi,\phi'}$;
5. $||\boldsymbol{K}_A\alpha||_{\phi,\phi'} = \textsc{Res}[||\alpha||_{\phi,\phi}, \phi]$;
6. $||\boldsymbol{K}_B\alpha||_{\phi,\phi'} = \textsc{Res}[||\alpha||_{\phi',\phi'}, \phi']$. $\blacksquare$

Intuitively, given an objective KB ϕ that A believes at all levels and an objective KB ϕ' that B believes at all levels, a conceptually simple reduction operator can be obtained. The reader may notice some similarity to the regression operator, *viz.* whenever $\boldsymbol{K}_A\alpha$ is encountered then the reduction is continued wrt the KB ϕ. Analogously, the reduction is continued wrt ϕ' whenever $\boldsymbol{K}_B\alpha$ is encountered.

We now present the main result for this section, by relating $\mathcal{R}$ and $\|\cdot\|_{\phi,\phi'}$:

Theorem 26 *Let Υ, Σ and Σ' be basic action theories. Suppose α is a basic bounded sentence of maximal A, B-depth k, j, then*

$$\Upsilon \wedge \psi \models \alpha \quad \textit{iff} \quad \models \Upsilon_0 \supset \|\mathcal{R}[\langle\rangle, \alpha]\|_{\Sigma_0, \Sigma_0'}.$$

where $\psi = OKnow_{\Sigma}[A, k] \wedge OKnow_{\Sigma'}[B, j]$.

That is, a query α perhaps with action operators is entailed by the background theory iff the regressed query reduced by the representation theorem wrt Σ_0 and Σ_0' is entailed by the set of sentences that are true initially. Thus, no modal reasoning is necessary. The proof for this theorem appears in Appendix B.

Example 27 Let us consider a projection query from Example 22. Consider, for example, the question whether (4) from Section 3 entails:

$$[pick_A(card_A)][see_A(card_A)]\boldsymbol{K}_A(num(card_A) = {}^{\#}1). \tag{9}$$

By means of Theorem 26, we get that $\models$ (4) $\supset$ (9) iff

$$\models \{num(card_A) = {}^{\#}1, num(card_B) = {}^{\#}52, \Sigma_0\} \supset \|\mathcal{R}[\Upsilon, \Sigma, \Sigma, \langle\rangle, (9)]\|_{\Sigma_0, \Sigma_0} \tag{10}$$

So for (10) to be true, first consider that $\mathcal{R}[\Upsilon, \Sigma, \Sigma, \langle\rangle, (9)]$ simplifies to

$$\exists x.\ x = num(card_A) \wedge \exists y.\ y = \mathrm{NIL} \wedge \boldsymbol{K}_A\alpha \tag{11}$$

where α is

$$y = \mathrm{NIL} \supset (num(card_A) = x \supset num(card_A) = {}^{\#}1).$$

Next, observe that $\|(11)\|_{\Sigma_0, \Sigma_0}$ yields

$$\exists x.\ x = num(card_A) \wedge \exists y.\ y = \mathrm{NIL} \wedge \textsc{Res}[\|\alpha\|_{\Sigma_0, \Sigma_0}, \Sigma_0]. \tag{12}$$

We then note that $\textsc{Res}[\|\alpha\|_{\Sigma_0, \Sigma_0}, \Sigma_0]$ reduces to

$$x = {}^{\#}1 \wedge y = \mathrm{NIL}. \tag{13}$$

The reduction is as follows. $\textsc{Res}[\|\alpha\|_{\Sigma_0, \Sigma_0}, \Sigma_0] = \textsc{Res}[\alpha, \Sigma_0]$ because α is objective. Now, $\textsc{Res}[\alpha, \Sigma_0]$ has 2 free variables: x and y. By the definition of $\textsc{Res}$, all possible substitutions n and m are chosen for x and y respectively from the names in $\Sigma_0 \cup \{\alpha\}$, to check whether $\textsc{Res}[\alpha^{x\ y}_{n\ m}, \Sigma_0]$ is true. But that is the case only for substitutions $^{\#}1$ and NIL for x and y respectively. Therefore, $\textsc{Res}[\alpha, \Sigma_0]$ yields (13). Replacing (13) in (12), we get:

$$\exists x[x = num(card_A) \wedge \exists y[y = \text{NIL} \wedge (x = {}^{\#}1 \wedge y = \text{NIL})]]. \tag{14}$$

Thus, from (10), we ask is it true that the following first-order formula is valid:

$$\{num(card_A) = {}^{\#}1, num(card_B) = {}^{\#}52, \Sigma_0\} \supset (14).$$

The answer is clearly *yes*, and therefore, $\models (4) \supset (9)$. ∎

So standard first-order theorem proving can be employed for reasoning about multiagent systems in $\mathcal{ES}_n$. There is a caveat, however. Unlike standard theorem proving, the set of basic bounded formulas that follow from a basic action theory by applying the representation theorem is not *recursively enumerable* (Rogers Jr., 1987). More precisely, in item 1 of $\textsc{Res}$'s definition, note that we appeal to validity, when returning TRUE, and appeal to falsifiability, when returning FALSE (Levesque & Lakemeyer, 2001).

4.4 Discussions

Before wrapping up the section, let us reflect on the limitations of the regression property and the representation theorem. Clearly, they represent a very special case, one of the form (say, for theories of depth 2):

$$\boldsymbol{O}_A(\phi \wedge \boldsymbol{O}_B\phi) \wedge \boldsymbol{O}_B(\psi \wedge \boldsymbol{O}_A\psi)$$

where if an agent only knows ϕ, she also believes other agents hold the same beliefs. It is not hard to generalize both the results to cases of the form:

$$\boldsymbol{O}_A(\phi \wedge \boldsymbol{O}_B\phi') \wedge \boldsymbol{O}_B(\psi \wedge \boldsymbol{O}_A\psi')$$

where, ϕ, ϕ', ψ and ψ' may differ arbitrarily. The idea, not surprisingly, is to relate the depth of the formula to the sentence believed by the agent at a corresponding depth. For example, if p is an atom, one would evaluate $[r]\boldsymbol{K}_A p$ wrt ϕ and $[r]\boldsymbol{K}_B p$ wrt ψ. At the next level, given a formula $[r]\boldsymbol{K}_A([r']\boldsymbol{K}_B p)$, we would evaluate $[r']\boldsymbol{K}_B p$ wrt ϕ', and given the resultant formula α, we would evaluate $[r]\boldsymbol{K}_A\alpha$ wrt ϕ. More precisely, the regression operator and the representation theorem will now be defined in terms of the sentences true in the real world, as well as the ones believed: ϕ, ϕ', ψ and ψ'. (That is, instead of three theories wrt which $\mathcal{R}$ is defined for knowledge bases of arbitrary depths, as in Section 4, we would have at most 5 theories for knowledge bases of depth 2.) Using the techniques presented here, it is then not hard to show that, yet again, we would obtain a property analogous to Theorem 26, where no modal reasoning will be necessary. Because we allow beliefs to differ arbitrarily after actions, we think expecting an initial specification that makes assumptions about what agents only know is reasonable. (Note also that ϕ and ϕ' can be any first-order theory, that is, no complete knowledge assumption is made here or anywhere else in this paper.) Therefore, this more general setting would cover a broad range of application domains. In return, only a slightly involved definition for $\mathcal{R}$ and $||\cdot||$ is needed.

However, there may be domains that make a case for still other kinds of initial states, such as

$$\boldsymbol{O}_A(\phi \wedge (\boldsymbol{K}_B\psi \vee \boldsymbol{K}_B\psi')) \wedge \boldsymbol{K}_B(\psi \wedge \boldsymbol{K}_A(\neg\phi))$$

is one where A only knows that B knows ψ or B knows ψ'. It is also an example where the modeler has not given a full characterization of B's knowledge base. Appealing to the underlying semantics to reason about properties of knowledge in such examples is well-defined, of course, since we are simply checking the validity of well-formed formulas in the logic. As far as the effectiveness of reasoning is concerned, note that most significantly, regression of basic bounded formulas is not limited to the nature of the initial theory. So this aspect is not the problem. But we have very little to say about the reduction of knowledge operators in these cases. Indeed, while a version of Theorem 21 is provable, where after regression, one would replace basic action theories by their initial components only, Theorem 26 need not hold. Thus, in such cases, modal reasoning will perhaps be necessary.

5. Related Work

This article focused on only knowing and knowledge in a multiagent dynamic setting. In particular, the modal dialect $\mathcal{ES}$ of the situation calculus, along with associated reduction theorems, were generalized to the many agent case.

The underlying language of the situation calculus has received a lot of attention in the action community. There are, of course, alternate formalisms, such as the fluent calculus (Thielscher, 1999) and other closely related approaches, such as those based on dynamic logics (Gerbrandy & Groeneveld, 1997; Demolombe, 2003; Demolombe, Herzig, & Varzinczak, 2003; Van Ditmarsch, Herzig, & De Lima, 2007). In particular, the action modality of $\mathcal{ES}$, which we inherit here, is taken from dynamic logic. However, there are significant differences. For example, Van Ditmarsch et al. (2007) consider an epistemic extension to dynamic logic with a regression property, but they are propositional, and do not consider only knowing. Demolombe (2003), on the other hand, considers a form of only knowing, but his work, like the original $\mathcal{ES}$, is restricted to the single agent case. Also, there is no notion of regression. For more details on how the single agent version of $\mathcal{ES}$ is related to various other action proposals, see the discussion by Lakemeyer and Levesque (2004). It is worth mentioning that the situation calculus itself has been previously extended to deal with multiple agents (Shapiro et al., 2002). Recently, in fact, Kelly and Pearce (2008) formulate the evaluation of epistemic queries, including queries about common knowledge (Fagin et al., 1995), by means of a meta-level operator using regression. In contrast to these ideas, we are mainly concerned with identifying how regression works in the presence of multiagent only knowing operators. As we have argued earlier, by being able to define initial knowledge in terms of what is only known one obtains a natural means of reasoning about both beliefs and non-beliefs. Moreover, the epistemic situation calculus of Scherl and Levesque does not have an equivalent of the representation theorem. Therefore, the other approaches would require a form of modal reasoning about the initial situation.[14] In other aspects, moreover, the approaches are not comparable. On the one hand, in contrast to Kelly and Pearce we observed in Section 2.2 that common knowledge cannot be captured with our semantics. On the other hand, integrating only knowing in the situation calculus when situation terms are explicit, as in the above derivatives of the Scherl-Levesque scheme, is problematic (Lakemeyer, 1996; Lakemeyer & Levesque, 1998, 2004).

14. Special cases for the reduction of knowledge are treated, for example, by Reiter (2001) and Lakemeyer and Lespérance (2012).

We note that reasoning about knowledge in multiagent systems is an important area in artificial intelligence, and numerous formal systems have been studied (Fagin et al., 1995; Wooldridge, 2009). For example, properties similar to ones obtained for the card game studied in this article are also considered in the work of Van Ditmarsch (2002). However, many of these systems are propositional. Most significantly, only knowing, and the feature that appropriate beliefs are entailments of a knowledge base that is only known, are not addressed.

As we pointed out, Levesque (1990) was among the first to propose a notion of only knowing in the logic $\mathcal{OL}$, but there are a number of related notions (Halpern & Moses, 1984; Hoek & Thijsse, 2002; Pratt-Hartmann, 2000). We do not discuss them here, nor their relationships to $\mathcal{OL}$, which is treated elsewhere (Rosati, 2000; Halpern & Lakemeyer, 2001; Levesque & Lakemeyer, 2001). Readers are also referred to the work of Levesque and Lakemeyer (2001) for a more comprehensive study on $\mathcal{OL}$; for example, it is shown that the compactness property does not hold for the objective fragment of $\mathcal{OL}$. Halpern and Pass (2009) consider a related probabilistic variant of only knowing for studying certain kinds of strategies in game theory.

Since Levesque's proposal, generalizations to the many agent case has been attempted in a number of papers (Lakemeyer, 1993; Halpern, 1993; Halpern & Lakemeyer, 2001; Waaler & Solhaug, 2005). But as we point out in earlier work (Belle & Lakemeyer, 2010a), these approaches have undesirable features. Our k-structures approach (Belle & Lakemeyer, 2010a), on the other hand, was shown to satisfactorily capture multiagent only knowing. In that work, we also discuss a number of other aspects of multiagent only knowing, including, for example, a sound and complete axiomatization for the propositional case.[15]

Finally, we remark that the intuition of k-structures seems closely related to the proposal of *knowledge structures* (Fagin, Halpern, & Vardi, 1991). Although restricted to a propositional language, and although actions and only knowing are not considered, the proposal is also based on epistemic states at various depths. (See Kaneko and Suzuki, 2003, for similar semantical notions in game theory.) For an agent, a *1-world* is simply a set of truth assignments to primitive propositions. This roughly corresponds to a set of worlds, similar to a 1-structure. However, a *2-world* considers the triple: truth assignment to primitive propositions, a 1-world for A, and a 1-world for B. So this differs from our proposal slightly. They also expect k-worlds to satisfy various constraints, including one about knowledge always being correct. Despite these differences, they are also motivated by the ease of capturing non-beliefs, and so an investigation on the correspondences between the two approaches is perhaps worthy of study.

6. Conclusions

This work considered reasoning about only knowing with many agents in dynamic domains. The language introduced is a first-order formalism that allows us to reason about knowledge, only knowing, actions and sensing. Only knowing has distinctive advantages from the view of a knowledge-based agent where it is possible to specify the sentences that precisely characterize a knowledge base, and then logically infer corresponding beliefs and non-beliefs (with quantifying-in) from that characterization. Building on previous work on multiagent only knowing (Belle & Lakemeyer, 2010a) and a modal fragment of the situation calculus (Lakemeyer & Levesque, 2004), a semantical

15. $\mathcal{OL}$'s axiomatization for the first-order language (Levesque, 1990) was shown to be incomplete by Halpern and Lakemeyer (1995); they also show that any complete axiomatization cannot be *recursive*.

account was first discussed. We showed that knowledge has appropriate properties, despite the semantics slightly deviating from the usual Kripke-structure account. We then considered the notion of a basic action theory, and explored projection tasks in terms of a simple card game. In particular, non-trivial knowledge change mechanisms after sensing was demonstrated in the formalism.

One important methodology for reasoning about actions in the literature is regression, extensively used in planning methodologies (Fritz, 2009), and we proved a version of regression for the formalism. From this, reasoning about actions and knowledge reduces to reasoning about knowledge in the initial state only. Next, we generalized the *representation theorem* (Levesque & Lakemeyer, 2001) to further reduce reasoning about knowledge in the initial state to first-order reasoning. Thus, no modal reasoning is necessary. We believe these results together with the underlying logic enhance the current paradigms for the logical modeling of intelligent agents (Fagin et al., 1995; Wooldridge, 2009), especially in the sense of formal specifications for knowledge-based systems.

There are many avenues for future work. An important observation on $\mathcal{OL}$ by Levesque (1990) is that when the knowledge base includes beliefs about itself, a certain flavor of nonmonotonicity is exhibited. In fact, the beliefs that logically follow can be related in a precise way to the fixed-point definition of *autoepistemic logic* (Moore, 1985). We have previously shown (Belle & Lakemeyer, 2010a) that these notions generalize to the multiagent case as well, which can be used for multiagent autoepistemic reasoning. For example, if Fred tells Sara that he recently bought a bird, Fred might come to assume that Sara believes then that the bird flies, without him explicitly suggesting such a fact. $\mathcal{ES}_n$, of course, would further allow these notions to be studied in a dynamic setting (Kakas, Michael, & Miller, 2008; Lakemeyer & Levesque, 2009).

For long-lived agents, regression would become infeasible after millions of actions, and so we would need to periodically update the knowledge base, which is referred to as *progression* (Lin & Reiter, 1997). STRIPS technology, for instance, is a simple form of progression (Reiter, 2001). Recently, the computational methodology of progression has been studied in the context of only knowing (Lakemeyer & Levesque, 2009). The idea, roughly, is that if the agent only knows a basic action theory $\Sigma_0 \cup \Sigma_{pre} \cup \Sigma_{post} \cup \Sigma_{sense}$, then after an action, the agent only knows another basic action theory $\Sigma_0' \cup \Sigma_{pre} \cup \Sigma_{post} \cup \Sigma_{sense}$, where Σ_0' is the progression of Σ_0. Here, only knowing characterizes the knowledge base in a precise way after doing actions. Our account might suggest ways to study these notions in a multiagent setting, where after actions, an agent would update not only her beliefs about the world but would also update her beliefs about what other agents know.

Finally, extensions for probabilistic nondeterminism (Gabaldon & Lakemeyer, 2007; Belle & Lakemeyer, 2011) and the development of *strategies* and *coalitions* between agents (Alur, Henzinger, & Kupferman, 2002; Giacomo, Lespérance, & Pearce, 2010) are worth exploring in an only knowing framework, perhaps along the lines of Halpern and Pass (2009).

Acknowledgements

We thank the reviewers for many helpful comments and suggestions. This work was carried out when the first author was supported by the graduate school GK 643 at RWTH Aachen University and funded by a DFG (German Research Foundation) scholarship.

Appendix A. Proof of Regression Property

In this section, we prove Theorem 21. We begin with a few useful lemmas before turning to the main theorem. In what follows, we will make use of the following special construction. Given a world w, we define another world w_Σ which is like w except that it satisfies $\Sigma_{pre}, \Sigma_{post}$ and Σ_{sense} sentences of Σ.

Definition 28 Let w be a world, $z \in \mathcal{Z}$ and Σ a basic action theory over fluents $\mathcal{F}$. Then w_Σ is a world satisfying the following conditions:

1. for $f \notin \mathcal{F}$, $w_\Sigma[f(n_1, \ldots, n_k), z] = w[f(n_1, \ldots, n_k), z]$;
2. for $f \in \mathcal{F}$, w_Σ is defined inductively by:
   (a) $w_\Sigma[f(n_1, \ldots, n_k), \langle\rangle] = w[f(n_1, \ldots, n_k), \langle\rangle]$;
   (b) $w_\Sigma[f(n_1, \ldots, n_k), z \cdot r] = m$ iff $w_\Sigma, z \models (\gamma_f)^{a\ y\ x_1,\ldots,x_k}_{r\ m\ n_1,\ldots,n_k}$;
3. $w_\Sigma[Poss(r), z] = 1$ iff $w_\Sigma, z \models \pi^a_r$;
4. $w_\Sigma[SF_i(r), z] = m$ iff $w_\Sigma, z \models \varphi_i{}^{a\ x}_{r\ m}$;

where γ_f, π and φ are the *rhs* of the successor state, precondition and sensing axioms respectively, appearing in the basic action theory Σ. ∎

The following properties can be shown regarding w_Σ in relation to w:

Lemma 29 *(Lakemeyer & Levesque, 2004)*

1. *For any w, w_Σ exists and is unique.*
2. *If $w \models \Sigma_0$ then $w_\Sigma \models \Sigma$.*
3. *If $w \models \Sigma$ then $w = w_\Sigma$.*
4. *Let α be any bounded objective sentence, and suppose that it is rectified and in* NF. *Let $z \in \mathcal{Z}$. Then $w \models \mathcal{R}[z, \alpha]$ iff $w_\Sigma, z \models \alpha$.*

Proof: The proof for the lemma is given elsewhere (Lakemeyer & Levesque, 2004). Later in this section, arguments analogous to their proof for item 4 will be needed, so we include the proof for this item here.

Item 4 is proven by induction on the length of α. We treat the length of $Poss(r)$ and $SF_i(r)$ as the length of π^a_r and $\varphi_i{}^a_r$ plus 1. We only consider the non-trivial cases below:

case $Poss(r)$.

We have $w_\Sigma, z \models Poss(r)$

iff $w_\Sigma, z \models \pi^a_r$ by definition of w_Σ

iff $w \models \mathcal{R}[z, \pi^a_r]$ by induction iff $w \models \mathcal{R}[z, Poss(r)]$ by definition of $\mathcal{R}$.

case $SF_i(r) = m$.

We have $w_\Sigma, z \models SF_i(r) = m$

iff $w_\Sigma, z \models \varphi_{i}{}^{a}_{r}$ by definition of w_Σ

iff $w \models \mathcal{R}[z, \varphi_{i}{}^{a}_{r}]$ by induction

iff $w \models \mathcal{R}[z, SF_i(r) = m]$ by definition of $\mathcal{R}$.

case fluents $f \in \mathcal{F}$. Note that, by definition of NF, ground atoms are of the form $f(n_1, \ldots, n_k) = m$. The proof is by sub-induction on z.

1. $w_\Sigma \models f(n_1, \ldots, n_k) = m$

   iff $w \models f(n_1, \ldots, n_k) = m$ by definition of w_Σ

   iff $w \models \mathcal{R}[\langle\rangle, f(n_1, \ldots, n_k) = m]$ by definition of $\mathcal{R}$.

2. $w_\Sigma, z \cdot r \models f(n_1, \ldots, n_k) = m$

   iff $w_\Sigma, z \models \gamma_{f}{}^{a}_{r}{}^{y}_{m}{}^{x_1,\ldots,x_k}_{n_1,\ldots,n_k}$ by definition of w_Σ

   iff $w \models \mathcal{R}[z, \gamma_{f}{}^{a}_{r}{}^{y}_{m}{}^{x_1,\ldots,x_k}_{n_1,\ldots,n_k}]$ by sub-induction

   iff $w \models \mathcal{R}[z \cdot r, f(n_1, \ldots, n_k) = m]$ by definition of $\mathcal{R}$. ∎

We now proceed to prove similar properties about epistemic states. Given e^k and a basic action theory Σ, let us define ${e_\Sigma}^k$ inductively by:

1. ${e_\Sigma}^1 = \{(w_\Sigma, \{\}) \mid (w, \{\}) \in e^1\}$;
2. ${e_\Sigma}^k = \{(w_\Sigma, {e_\Sigma}^{k-1}) \mid (w, e^{k-1}) \in e^k\}$.

In addition, using our notation for nested only knowing operators (see Section 3), for brevity, let

- $\psi_0 = OKnow_{\Sigma_0}[A, k] \wedge OKnow_{\Sigma_0{}'}[B, j]$, and
- $\psi = OKnow_{\Sigma}[A, k] \wedge OKnow_{\Sigma'}[B, j]$.

Then, item 2 of Lemma 29 is extended for knowledge in the following manner:

Lemma 30 *Suppose $e^k_A, e^j_B, w \models \psi_0$. Then ${e_\Sigma}^k_A, {e_{\Sigma'}}^j_B, w \models \psi$.*

Proof: The proof is a simple induction on the modal depth (Definition 13) of the background theory. Recall that when the modal depth of the background theory is l, then we have a sentence of the form $OKnow_{\Sigma_0}[A, k] \wedge OKnow_{\Sigma_0{}'}[B, j]$ such that $k \leq l$, $j \leq l$ and k or j is l.

The base case is a theory of modal depth 1. So suppose $e^1_A, e^1_B, w \models \boldsymbol{O}_A(\Sigma_0) \wedge \boldsymbol{O}_B(\Sigma_0{}')$. We will now show $(w', \{\}) \in {e_\Sigma}^1_A$ iff $w' \models \Sigma$. From that, we get ${e_\Sigma}^1_A, \{\}, w \models \boldsymbol{O}_A\Sigma$. The case of ${e_{\Sigma'}}^1_B$ is entirely analogous, by means of which we have shown ${e_\Sigma}^1_A, {e_{\Sigma'}}^1_B, w \models \boldsymbol{O}_A(\Sigma) \wedge \boldsymbol{O}_B(\Sigma')$.

Suppose $w \models \Sigma$. Then $w \models \Sigma_0$ and therefore, by assumption, $(w, \{\}) \in e^1_A$. By Lemma 29, $w = w_\Sigma$ and therefore, $(w, \{\}) \in {e_\Sigma}^1_A$.

Conversely, let $(w, \{\}) \in e_{\Sigma}{}^{1}_{A}$. By definition, there is a $(w', \{\}) \in e^1_A$ such that $w'_{\Sigma} = w$. But since $w' \models \Sigma_0$, it follows from Lemma 29 that $w \models \Sigma$. Thus, $e_{\Sigma}{}^{1}_{A}, \{\}, w \models \boldsymbol{O}_A(\Sigma)$.

Assume that the hypothesis holds for theories of modal depth $k-1$, that is, if e^{k-1}_A satisfies $OKnow_{\Sigma_0}[A, k-1]$ then $e_{\Sigma}{}^{k}_{A}$ satisfies $OKnow_{\Sigma}[A, k-1]$, and similarly for B. Now, suppose $e^k_A, e^j_B, w \models \psi_0$. Then, $(w', e^{k-1}_B) \in e^k_A$ iff $e^k_A, e^{k-1}_B, w' \models \Sigma_0 \wedge OKnow_{\Sigma_0}[B, k-1]$. We now show $(w', e^{k-1}_B) \in e_{\Sigma}{}^{k}_{A}$ iff $e_{\Sigma}{}^{k}_{A}, e^{k-1}_B, w' \models \Sigma \wedge OKnow_{\Sigma}[B, k-1]$, from which we get $e_{\Sigma}{}^{k}_{A}, \{\}, w \models OKnow_{\Sigma}[A, k]$. The argument is symmetric for e^j_B, and therefore, the lemma's claim follows.

Consider any e^{k-1}_B and w such that $e_{\Sigma}{}^{k}_{A}, e^{k-1}_B, w \models \Sigma \wedge OKnow_{\Sigma}[B, k-1]$. Now, consider $e'^{\,k-1}_B$ such that $\{\}, e'^{\,k-1}_B, w \models OKnow_{\Sigma_0}[B, k-1]$. Since $w \models \Sigma$, by Lemma 29 $w = w_{\Sigma}$ and also, $w \models \Sigma_0$. It follows that $(w, e'^{\,k-1}_B) \in e^k_A$ by assumption. By induction hypothesis, $\{\}, e_{\Sigma}{}'^{k-1}_{B}, w \models OKnow_{\Sigma}[B, k-1]$. By definition, $(w, e_{\Sigma}{}'^{k-1}_{B}) \in e_{\Sigma}{}^{k}_{A}$. An easy argument shows that $e_{\Sigma}{}'^{k-1}_{B} = e^{k-1}_B$.

Conversely, consider any $(w, e^{k-1}_B) \in e^k_A$. By assumption, $\{\}, e^{k-1}_B, w \models \Sigma_0 \wedge OKnow_{\Sigma_0}[B, k-1]$. By Lemma 14, $w_{\Sigma} \models \Sigma$. By induction hypothesis, $\{\}, e_{\Sigma}{}^{k-1}_{B}, w \models OKnow_{\Sigma}[B, k-1]$. By definition, $(w_{\Sigma}, e_{\Sigma}{}^{k-1}_{B}) \in e_{\Sigma}{}^{k}_{A}$. ∎

We now generalize item 4 of Lemma 29 for knowledge.

Lemma 31 $e^k_A, e^j_B, w \models \mathcal{R}[\Upsilon, \Sigma, \Sigma', z, \alpha]$ *iff* $e_{\Sigma}{}^{k}_{A}, e_{\Sigma'}{}^{j}_{B}, w_{\Upsilon}, z \models \alpha$.

Proof: The proof is by induction on z, and a sub-induction on α.

Let $z = \langle\rangle$. The case of objective formulas proceeds exactly as in Lemma 29. So let us consider the case of A-subjective formulas.

We have $e_{\Sigma}{}^{k}_{A}, e_{\Sigma'}{}^{j}_{B}, w_{\Upsilon}, z \models \boldsymbol{K}_A\alpha$

iff for all $(w, e^{k-1}_B) \in e_{\Sigma}{}^{k}_{A}$, $e_{\Sigma}{}^{k}_{A}, e^{k-1}_B, w \models \alpha$

iff for all $(w, e^{k-1}_B) \in e^k_A$, $e_{\Sigma}{}^{k}_{A}, e_{\Sigma}{}^{k-1}_{B}, w_{\Sigma} \models \alpha$ by definition of $e_{\Sigma}{}^{k}_{A}$

iff for all $(w, e^{k-1}_B) \in e^k_A$, $e^k_A, e^{k-1}_B, w \models \mathcal{R}[\langle\rangle, \alpha]$ by sub-induction

iff $e^k_A, e^j_B, w \models \boldsymbol{K}_A\mathcal{R}[\langle\rangle, \alpha]$

iff $e^k_A, e^j_B, w \models \mathcal{R}[\langle\rangle, \boldsymbol{K}_A\alpha]$ by definition of $\mathcal{R}$.

The case of B-subjective formulas is symmetric.

Now, we consider the case of $z \cdot r$. The proof is precisely as in the base case, except for subjective formulas, which we prove as follows. We show the argument for A-subjective formulas. The arguments for B-subjective formulas is symmetric.

$e_{\Sigma}{}^{k}_{A}, e_{\Sigma'}{}^{j}_{B}, w_{\Upsilon}, z \cdot r \models \boldsymbol{K}_A\alpha$

iff $e_{\Sigma}{}^{k}_{A}, e_{\Sigma'}{}^{j}_{B}, w_{\Upsilon}, z \models [r]\boldsymbol{K}_A\alpha$ by definition

iff $e_{\Sigma}{}^{k}_{A}, e_{\Sigma'}{}^{j}_{B}, w_{\Upsilon}, z \models \beta^a_r$ where β is the *rhs* of Theorem 19 for $[r]\boldsymbol{K}_A\alpha$

iff $e^k_A, e^j_B, w \models \mathcal{R}[z, \beta^a_r]$ by the main induction iff $e^k_A, e^j_B, w \models \mathcal{R}[z \cdot r, \boldsymbol{K}_A\alpha]$ by definition of $\mathcal{R}$. ∎

We are now ready to prove Theorem 21. We restate the claim below:

Theorem 21 *Suppose α is a bounded basic sentence of maximal A, B-depth k, j. Let Υ, Σ and Σ' be basic action theories. Then $\mathcal{R}[\langle\rangle, \alpha]$ is a static sentence and satisfies:*

$$\Upsilon \wedge \psi \models \alpha \quad \textit{iff} \quad \Upsilon_0 \wedge \psi_0 \models \mathcal{R}[\langle\rangle, \alpha]$$

where $\psi = \mathit{OKnow}_\Sigma[A, k] \wedge \mathit{OKnow}_{\Sigma'}[B, j]$
$\psi_0 = \mathit{OKnow}_{\Sigma_0}[A, k] \wedge \mathit{OKnow}_{\Sigma_0'}[B, j]$.

Proof: Let us denote $\Upsilon \wedge \psi$ as Γ and $\Upsilon_0 \wedge \psi_0$ as Γ_0.

For the only-if direction, suppose that $\Gamma \models \alpha$ and suppose that $e^k_A, e^j_B, w \models \Gamma_0$. That is, $w \models \Upsilon_0$ and by Lemma 29, $w_\Upsilon \models \Upsilon$. Further, by Lemma 30, $e_{\Sigma}{}^k_A, e_{\Sigma'}{}^j_B, w_\Upsilon \models \Gamma$. By assumption, $e_{\Sigma}{}^k_A, e_{\Sigma'}{}^j_B, w_\Upsilon \models \alpha$. Then, by Lemma 31, $e^k_A, e^j_B, w \models \mathcal{R}[\langle\rangle, \alpha]$.

Conversely, suppose that $\Gamma_0 \models \mathcal{R}[\langle\rangle, \alpha]$ and let $e^k_A, e^j_B, w \models \Gamma$. Then $w \models \Upsilon_0$. Suppose that $e'^k_A, e'^j_B, w \models \psi_0$. By assumption $e'^k_A, e'^j_B, w \models \mathcal{R}[\langle\rangle, \alpha]$. By Lemma 31, $e_{\Sigma}{}'^k_A, e_{\Sigma'}{}'^j_B, w_\Upsilon \models \alpha$. By Lemma 29, $w_\Upsilon = w$. By Lemma 30, $e_{\Sigma}{}'^k_A, e_{\Sigma'}{}'^j_B, w_\Upsilon \models \Gamma$. Since both e^k_A and $e_{\Sigma}{}'^k_A$ are k-structures for A where $\mathit{OKnow}_\Sigma[A, k]$ holds, a simple induction argument shows that $e_{\Sigma}{}'^k_A = e^k_A$. Analogously, $e_{\Sigma'}{}'^j_B$ and e^j_B are the same. Therefore $e^k_A, e^j_B, w \models \alpha$. ∎

Appendix B. Proof of Representation Theorem

In this section, we prove Theorem 26. We proceed first by relating valid fluent sentences in $\mathcal{ES}_n$ to its non-dynamic fragment $\mathcal{OL}_n$ (Belle & Lakemeyer, 2010a). For this, we will only go over a few essential details of $\mathcal{OL}_n$. Roughly speaking, $\mathcal{OL}_n$ is $\mathcal{ES}_n$ without the dynamic operators $\{[t], \Box\}$ and distinguished symbols $\{\mathit{Poss}, \mathit{SF}_i.\}$ A *static* world $w \in \mathcal{W}^*$ is any function from primitive sentences to $\{0, 1\}$ and from primitive terms to standard names. Epistemic states in $\mathcal{OL}_n$ are k-structures over such static worlds. All other notions carry over to $\mathcal{OL}_n$, by simply ignoring dynamic aspects. For example, we specify the semantics for $\boldsymbol{K}_A\alpha$ wrt the triple (e^k_A, e^j_B, w) for $w \in \mathcal{W}^*$ as follows:

- $e^k_A, e^j_B, w \models \boldsymbol{K}_A\alpha$ iff for all $w' \in \mathcal{W}^*$, for all e^{k-1}_B, if $(w', e^{k-1}_B) \in e^k_A$, then $e^k_A, e^{k-1}_B, w' \models \alpha$.

In other words, roughly, we dropped the action sequence z and the compatibility relation $\simeq^A_z$ from the semantical definition of $\boldsymbol{K}_A\alpha$ in $\mathcal{ES}_n$. Terminology for formulas, such as *objective* and *basic* is analogously defined for $\mathcal{OL}_n$.

We now present three formal properties regarding $\mathcal{OL}_n$ and $\mathcal{ES}_n$ sentences:

Lemma 32 *For any $\alpha \in \mathcal{OL}_n$, α is valid in $\mathcal{OL}_n$ iff α is valid in $\mathcal{ES}_n$.*

Lemma 33 *If $\alpha \in \mathcal{ES}_n$ is a fluent sentence and z is any action sequence, then $\mathcal{R}[\Upsilon, \Sigma, \Sigma', z, \alpha]$ is an objective $\mathcal{OL}_n$-sentence.*

Both proofs are straightforward generalizations of analogous results regarding $\mathcal{OL}$ and $\mathcal{ES}$, appearing as Theorem 6, Lemma 9 and Lemma 10 in the work of Lakemeyer and Levesque (2004), and therefore not reproduced here. For example, with Lemma 32, the main technical scheme is to relate $\mathcal{ES}$ (and thus $\mathcal{ES}_n$) worlds and $\mathcal{OL}$ (and thus $\mathcal{OL}_n$) worlds. For Lemma 33, clearly objective $\mathcal{OL}$ sentences are also objective $\mathcal{OL}_n$ sentences, and so the claim follows.

Lemma 34 *If α is a bounded basic sentence and z is any action sequence, then $\mathcal{R}[\Upsilon, \Sigma, \Sigma', z, \alpha]$ is a basic $\mathcal{OL}_n$ sentence.*

Proof: The proof is by induction on α. If α is a fluent sentence then the argument is immediate owing to Lemma 33. For *Poss*, $\mathcal{R}[z, Poss(t)] = \mathcal{R}[z, \pi^a_t]$, but π^a_t is a fluent formula, so Lemma 33 applies. For SF_i, $\mathcal{R}[z, SF_i(t) = t'] = \mathcal{R}[z, {\varphi_i}^{a\ y}_{t\ t'}]$, and again, ${\varphi_i}^{a\ y}_{t\ t'}$ is a fluent formula. For $[t]$, $R[z, [t]\alpha] = R[z \cdot t, \alpha]$ which is a basic $\mathcal{OL}_n$ sentence by induction.

For $\boldsymbol{K}_A$, we do a sub-induction on z. The case for $\boldsymbol{K}_B$ is analogous. $\mathcal{R}[\langle\rangle, \boldsymbol{K}_A\alpha] = \boldsymbol{K}_A\mathcal{R}[\langle\rangle, \alpha]$. Since by the main induction, $\mathcal{R}[\langle\rangle, \alpha]$ is basic, and so $\boldsymbol{K}_A\mathcal{R}[\langle\rangle, \alpha]$ is also basic. $\mathcal{R}[z \cdot t, \boldsymbol{K}_A\alpha] = \mathcal{R}[z, \beta^a_t]$, where β is the *rhs* of Theorem 19 for $[t]\boldsymbol{K}_A\alpha$. By the sub-induction hypothesis and Lemma 33, β^a_t is also basic. ∎

We will now prove three main results that are essential for Theorem 26. To prepare for that, given static worlds $\mathcal{W}^*$ and an objective $\mathcal{OL}_n$-sentence ϕ, let:

- $\mathcal{W}_\phi = \{w \mid w \models \phi, w \in \mathcal{W}^*\}$;
- ${e_\phi}^1 = \mathcal{W}_\phi \times \{\{\}\}$;
- ${e_\phi}^k = \{(w, {e_\phi}^{k-1}) \mid w \in \mathcal{W}_\phi\}$.

In the sequel, benefiting from Lemma 32, we simply argue using $\mathcal{OL}_n$-models, that is, by ignoring dynamic notions.

Lemma 35 *Let ϕ and ϕ' be objective $\mathcal{OL}_n$ sentences and let ${e_\phi}^k_A$ and ${e_{\phi'}}^j_B$ be as above. Let α be any objective formula with free variables $x_1, \ldots, x_k$. For any vector of standard names $n_1, \ldots, n_k$ and world w:*

$$ {e_\phi}^k_A, {e_{\phi'}}^j_B, w \models \boldsymbol{K}_A\alpha^{x_1,\ldots,x_k}_{n_1,\ldots,n_k} \;\; \textit{iff} \;\; \models \textsc{Res}[\alpha, \phi]^{x_1,\ldots,x_k}_{n_1,\ldots,n_k}. $$

Analogously for $\boldsymbol{K}_B\alpha^{x_1,\ldots,x_k}_{n_1,\ldots,n_k}$.

Proof: From Lemma 7, it follows that ${e_\phi}^k_A, \{\}, w \models \boldsymbol{K}_A\alpha^{x_1,\ldots,x_k}_{n_1,\ldots,n_k}$ iff ${e_\phi}_A\!\downarrow^k_1, \{\}, w \models \boldsymbol{K}_A\alpha^{x_1,\ldots,x_k}_{n_1,\ldots,n_k}$ because $\boldsymbol{K}_A\alpha$ has A-depth 1. So it is sufficient to show that:

$$ {e_\phi}_A\!\downarrow^k_1, \{\}, w \models \boldsymbol{K}_A\alpha^{x_1,\ldots,x_k}_{n_1,\ldots,n_k} \;\; \text{iff} \;\; \models \textsc{Res}[\alpha, \phi]^{x_1,\ldots,x_k}_{n_1,\ldots,n_k}. \tag{15} $$

Note that ${e_\phi}_A\!\downarrow^k_1 = \{(w, \{\}) \mid w \models \phi\}$, and so (15) can be simply proved in $\mathcal{OL}$ (Levesque & Lakemeyer, 2001, Lemma 7.2.2). ∎

Theorem 36 *Let α be any basic $\mathcal{OL}_n$ formula of maximal A, B-depth k, j and with free variables $x_1, \ldots, x_k$. Let ${e_\phi}^k_A, {e_{\phi'}}^j_B$ be as before, w any world, and $n_1, \ldots, n_k$ be a vector of names. Then*

$$ {e_\phi}^k_A, {e_{\phi'}}^j_B, w \models \alpha^{x_1,\ldots,x_k}_{n_1,\ldots,n_k} \; \textit{iff} \; w \models \|\alpha\|_{\phi,\phi'}{}^{x_1,\ldots,x_k}_{n_1,\ldots,n_k}. $$

Proof: The proof is by induction on the structure of α. If α is an atom or an equality, the lemma clearly holds since α is objective. By induction, the lemma also holds for negations, disjunctions and quantifiers.

Now, consider $\boldsymbol{K}_A\alpha$. (The case of $\boldsymbol{K}_B\alpha$ is symmetric.) We have

$e_{\phi}{}^k_A, \{\}, w \models \boldsymbol{K}_A\alpha^{x_1,\ldots,x_k}_{n_1,\ldots,n_k}$

iff $e_{\phi}{}^k_A, e^{k-1}_B, w' \models \alpha^{x_1,\ldots,x_k}_{n_1,\ldots,n_k}$ for all $(w', e^{k-1}_B) \in e_{\phi}{}^k_A$

iff $w' \models \|\alpha\|_{\phi,\phi'}{}^{x_1,\ldots,x_k}_{n_1,\ldots,n_k}$ by the induction hypothesis

iff $e_{\phi}{}^k_A, \{\}, w \models \boldsymbol{K}_A\|\alpha\|_{\phi,\phi'}{}^{x_1,\ldots,x_k}_{n_1,\ldots,n_k}$ since $\|\alpha\|_{\phi,\phi'}{}^{x_1,\ldots,x_k}_{n_1,\ldots,n_k}$ is objective

iff $\models \textsc{Res}[\|\alpha\|_{\phi,\phi'}{}^{x_1,\ldots,x_k}_{n_1,\ldots,n_k}, \phi]^{x_1,\ldots,x_k}_{n_1,\ldots,n_k}$ by Lemma 35

iff $\models \|\boldsymbol{K}_A\alpha\|_{\phi,\phi'}{}^{x_1,\ldots,x_k}_{n_1,\ldots,n_k}$ by definition of $\textsc{Res}$

iff $w \models \|\boldsymbol{K}_A\alpha\|_{\phi,\phi'}{}^{x_1,\ldots,x_k}_{n_1,\ldots,n_k}$ because the result of $\textsc{Res}$ is an objective formula that does not use predicates and function symbols. Therefore, $\|\boldsymbol{K}_A\alpha\|_{\phi,\phi'}{}^{x_1,\ldots,x_k}_{n_1,\ldots,n_k}$ is either valid or unsatisfiable. ∎

Theorem 37 *Suppose α is of maximal A, B-depth k, j. Let ϕ, ϕ' and θ be objective $\mathcal{OL}_n$ sentences. Then*

$$\theta \wedge \psi \models \alpha \textit{ iff } \models \theta \supset \|\alpha\|_{\phi,\phi'}.$$

where $\psi = OKnow_{\phi}[A, k] \wedge OKnow_{\phi'}[B, j]$.

Proof: For the if direction, suppose (e^k_A, e^j_B, w) is a model of $\psi \wedge \theta$. It is easy to verify that $e^k_A = e_{\phi}{}^k_A$ and $e^j_B = e_{\phi'}{}^j_B$, and so, w is any world satisfying θ. Since $\psi \wedge \theta \models \alpha$, $e^k_A, e^j_B, w \models \alpha$ iff $w \models \|\alpha\|_{\phi,\phi'}$ by Theorem 36. So any model of θ satisfies $\|\alpha\|_{\phi,\phi'}$. Therefore, $\theta \models \|\alpha\|_{\phi,\phi'}$ or $\models \theta \supset \|\alpha\|_{\phi,\phi'}$.

Conversely, suppose $\theta \models \|\alpha\|_{\phi,\phi'}$. Now, let (e^k_A, e^j_B, w) be any model of $\psi \wedge \theta$. It is easy to verify that $e^k_A = e_{\phi}{}^k_A$ and $e^j_B = e_{\phi'}{}^j_B$. Further, since $w \models \theta$ we have $w \models \|\alpha\|_{\phi,\phi'}$. By Theorem 36, $e^k_A, e^j_B, w \models \alpha$. ∎

Finally, we turn to the proof for Theorem 26. We restate the claim below.

Theorem 26 *Let Υ, Σ and Σ' be basic action theories. Suppose α is a basic bounded sentence of maximal A, B-depth k, j, then*

$$\Upsilon \wedge \psi \models \alpha \quad \textit{iff} \quad \models \Upsilon_0 \supset \|\mathcal{R}[\langle\rangle, \alpha]\|_{\Sigma_0,\Sigma_0'}.$$

where $\psi = OKnow_{\Sigma}[A, k] \wedge OKnow_{\Sigma'}[B, j]$.

Proof: We have $\Upsilon \wedge OKnow_{\Sigma}[A, k] \wedge OKnow_{\Sigma'}[B, j] \models \alpha$

iff $\Upsilon_0 \wedge OKnow_{\Sigma_0}[A, k] \wedge OKnow_{\Sigma_0'}[B, j] \models \mathcal{R}[\langle\rangle, \alpha]$ by the regression property Theorem 21

iff $\Upsilon_0 \wedge OKnow_{\Sigma_0}[A, k] \wedge OKnow_{\Sigma_0'}[B, j] \supset \mathcal{R}[\langle\rangle, \alpha]$ is valid in $\mathcal{OL}_n$ by Lemma 32, owing to the fact that $\mathcal{R}[\langle\rangle, \alpha]$ is also a (basic) $\mathcal{OL}_n$ sentence by Lemma 34

iff $\Upsilon_0 \supset \|\mathcal{R}[\langle\rangle, \alpha]\|_{\Sigma_0,\Sigma_0'}$ is valid in $\mathcal{OL}_n$ by Theorem 37. ∎

References

Alur, R., Henzinger, T. A., & Kupferman, O. (2002). Alternating-time temporal logic. *J. ACM*, *49*(5), 672–713.

Belle, V., & Lakemeyer, G. (2010a). Multi-agent only-knowing revisited. In *Proc. KR*, pp. 49–60.

Belle, V., & Lakemeyer, G. (2010b). Reasoning about imperfect information games in the epistemic situation calculus. In *Proc. AAAI*, pp. 255–261.

Belle, V., & Lakemeyer, G. (2011). A semantical account of progression in the presence of uncertainty. In *Proc. AAAI*, pp. 165–170.

Demolombe, R. (2003). Belief change: from situation calculus to modal logic. In *Proc. Nonmonotonic Reasoning, Action, and Change (NRAC)*.

Demolombe, R., Herzig, A., & Varzinczak, I. (2003). Regression in modal logic. *Journal of Applied Non-Classical Logics*, *13*(2), 165–185.

Fagin, R., Halpern, J. Y., Moses, Y., & Vardi, M. Y. (1995). *Reasoning About Knowledge*. MIT Press.

Fagin, R., Halpern, J. Y., & Vardi, M. Y. (1991). A model-theoretic analysis of knowledge. *J. ACM*, *38*(2), 382–428.

Fritz, C. (2009). *Monitoring the Generation and Execution of Optimal Plans*. Ph.D. thesis, University of Toronto.

Gabaldon, A., & Lakemeyer, G. (2007). ESP: A logic of only-knowing, noisy sensing and acting. In *Proc. AAAI*, pp. 974–979.

Gerbrandy, J., & Groeneveld, W. (1997). Reasoning about information change. *J. of Logic, Lang. and Inf.*, *6*(2), 147–169.

Giacomo, G. D., Lespérance, Y., & Pearce, A. R. (2010). Situation calculus based programs for representing and reasoning about game structures. In *KR*.

Halpern, J. Y. (1993). Reasoning about only knowing with many agents. In *Proc. AAAI*, pp. 655–661.

Halpern, J. Y., & Lakemeyer, G. (1995). Levesque's axiomatization of only knowing is incomplete. *Artificial Intelligence*, *74*(2), 381–387.

Halpern, J. Y., & Moses, Y. (1984). Towards a theory of knowledge and ignorance: Preliminary report. In *Proc. NMR*, pp. 125–143.

Halpern, J. Y., & Pass, R. (2009). A logical characterization of iterated admissibility. In *Proc. TARK*, pp. 146–155.

Halpern, J., & Lakemeyer, G. (2001). Multi-agent only knowing. *Journal of Logic and Computation*, *11*(1), 251–265.

Harel, D., Kozen, D., & Tiuryn, J. (2000). *Dynamic logic*. The MIT Press.

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

Hoek, W. V. D., & Thijsse, E. (2002). A general approach to multi-agent minimal knowledge: With tools and samples. *Studia Logica*, *72*(1), 61–84.

Kakas, A. C., Michael, L., & Miller, R. (2008). Fred meets tweety. In *ECAI*, pp. 747–748.

Kaneko, M., & Suzuki, N.-Y. (2003). Epistemic models of shallow depths and decision making in games: Horticulture. *The Journal of Symbolic Logic*, *68*(1), pp. 163–186.

Kaplan, D. (1968). Quantifying in. *Synthese*, *19*(1), 178–214.

Kelly, R. F., & Pearce, A. R. (2008). Complex epistemic modalities in the situation calculus. In *Proc. KR*, pp. 611–620.

Kripke, S. (1963). Semantical considerations on modal logic. *Acta Philosophica Fennica*, *16*, 83–94.

Lakemeyer, G. (1996). Only knowing in the situation calculus. In *Proc. KR*, pp. 14–25.

Lakemeyer, G., & Levesque, H. J. (2011). A semantic characterization of a useful fragment of the situation calculus with knowledge. *Artificial Intelligence*, *175*, 142–164.

Lakemeyer, G., & Levesque, H. J. (2004). Situations, si! situation terms, no!. In *Proc. KR*, pp. 516–526.

Lakemeyer, G., & Levesque, H. (1998). AOL: a logic of acting, sensing, knowing, and only knowing. In *Proc. KR*, pp. 316–329.

Lakemeyer, G. (1993). All they know: A study in multi-agent autoepistemic reasoning. In *Proc. IJCAI*, pp. 376–381.

Lakemeyer, G., & Lespérance, Y. (2012). Efficient reasoning in multiagent epistemic logics. In *Proc. ECAI*, pp. 498–503.

Lakemeyer, G., & Levesque, H. (2009). A semantical account of progression in the presence of defaults. In *Conceptual Modeling: Foundations and Applications*, pp. 82–98. Springer.

Levesque, H. J. (1990). All I know: a study in autoepistemic logic. *Artificial Intelligence*, *42*(2-3), 263–309.

Levesque, H., & Lakemeyer, G. (2001). *The logic of knowledge bases*. The MIT Press.

Lin, F., & Reiter, R. (1997). How to progress a database. *Artificial Intelligence*, *92*(1-2), 131–167.

McCarthy, J., & Hayes, P. J. (1969). Some philosophical problems from the standpoint of artificial intelligence. In *Machine Intelligence*, pp. 463–502.

Moore, R. C. (1985). Semantical considerations on nonmonotonic logic. *Artificial Intelligence*, *25*(1), 75–94.

Pratt-Hartmann, I. (2000). Total knowledge. In *Proc. AAAI*, pp. 423–428.

Reiter, R. (2001). *Knowledge in action: logical foundations for specifying and implementing dynamical systems*. MIT Press.

Rogers Jr., H. (1987). *Theory of recursive functions and effective computability*. The MIT Press.

Rosati, R. (2000). On the decidability and complexity of reasoning about only knowing. *Artificial Intelligence*, *116*(1-2), 193–215.

Scherl, R. B., & Levesque, H. J. (2003). Knowledge, action, and the frame problem. *Artificial Intelligence*, *144*(1-2), 1–39.

Shapiro, S., Lespérance, Y., & Levesque, H. (2002). The cognitive agents specification language and verification environment for multiagent systems. In *Proc. AAMAS*, pp. 19–26.

Thielscher, M. (1999). From situation calculus to fluent calculus: state update axioms as a solution to the inferential frame problem. *Artificial Intelligence*, *111*(1-2), 277–299.

Van Ditmarsch, H., Herzig, A., & De Lima, T. (2007). Optimal regression for reasoning about knowledge and actions. In *Proc. AAAI*, pp. 1070–1075.

Van Ditmarsch, H. (2002). Descriptions of game actions. *Journal of Logic, Language and Information*, *11*(3), 349–365.

Waaler, A., & Solhaug, B. (2005). Semantics for multi-agent only knowing: extended abstract. In *Proc. TARK*, pp. 109–125.

Wooldridge, M. (2009). *An Introduction to Multiagent Systems* (2 edition). Wiley, Chichester, UK.