Rule-based Reasoners in Epistemic Logic

Anthia Solaki

ILLC, University of Amsterdam
a.solaki2@uva.nl

Abstract. In this paper, we offer a balanced response to the problem of logical omniscience, whereby agents are modeled as non-omniscient yet still logically competent reasoners. To achieve this, we account for the deductive steps that form the epistemic state of an agent. In particular, we introduce operators for applications of inference rules and design a possible-worlds model which is (a) equipped with a syntactic valuation, determining the agent's (explicit) knowledge, and (b) suitably structured by rule-induced transitions between worlds. As a result, we obtain a detailed analysis of the agent's reasoning processes. We then offer validities that exemplify how the problem of logical omniscience is avoided and compare our response to others in the literature. A sound and complete axiomatization is also provided. We finally show how simple extensions of this setting make it compatible with tools from Dynamic Epistemic Logic (DEL) and open to the incorporation of empirical findings on human reasoning.

Keywords: Rule-based reasoners. Epistemic logic. Dynamic epistemic logic. Logical omniscience. Bounded reasoning. Resource-bounded agents. Minimal rationality. Human reasoning.

1 Introduction

Standard (S5) epistemic logic, using possible-worlds semantics, suffers from the *problem of logical omniscience* [13]: agents are modelled as reasoners with unlimited inferential power, always knowing whatever follows logically from what they know. This stark contrast with reality is also witnessed by experimental results indicating that subjects are systematically fallible in reasoning tasks [21, 22]. It is even from a normative view that the standard account is insufficient, for it disregards the underlying reasoning of the agent and thus the restrictions on what can be *feasibly* asked of her. Therefore, knowledge should not be subject to logical closure principles. This, however, need not entail that agents are logically incompetent. While we often fail in complex inferences (e.g. due to lack of resources), we do engage in bounded reasoning: knowing that it is raining, and that we need a raincoat whenever it is raining, we do take a raincoat before leaving home. The empirical data also contributes to the case for logical competence, and as proposed in [9], we should seek a standard of *Minimal Rationality*. Drawing on these, we aim at modelling how an agent should *come to know* whatever can be feasibly reached from her epistemic state.

In the twofold project of modelling a non-omniscient yet competent agent, we take on board the observations found in [7]. The deductive steps underpinning knowledge should be clearly reflected in an epistemic framework and this should still be compatible with "external" informational acts, as studied in DEL. We also place another desideratum: in principle, we should be able to employ empirical facts provided by cognitive scientists.

While many attempts have dealt with logical omniscience, not every attempt pursues a solution along the lines just described. Rule-based approaches, mainly applied on Artificial Intelligence, have paved the way towards our direction. Konolige [16] uses *belief sets* closed under an (incomplete) set of inference rules, but such (weaker) closure properties do not suffice to capture the agent's reasoning nor its cognitive load. Similar remarks apply to attempts which use modalities for reasoning processes [11], state-transitions due to inference [2, 3, 4], or arbitrary rule applications [14]. Collapsing reasoning processes to a modality, without a detailed analysis of their composition, would not help us determine what eventually makes them halt nor exploit investigations in psychology of reasoning which usually study *individual* inference rules on the grounds of cognitive difficulty. Interestingly, in [17], the author develops a logic where rules, accompanied by cognitive costs, are explicitly introduced in the language, but he gives no semantics, rendering the effect of his rule-operators unclear and the choice of axioms controversial.[1] Awareness settings [12] discern implicit and explicit attitudes, avoiding omniscience with respect to the latter, which additionally ask that agents are *aware* of a formula. Yet, an arbitrary syntactic awareness-filter cannot be associated with logical competence, and even if ad-hoc modifications are imposed (e.g. awareness closure under subformulas), forms of the problem are retained.[2]

The remainder is organized as follows: we first present our basic setting and explain how it contributes to the solution of the problem (Section 2). We then give a sound and complete axiomatization in Section 3 and in Section 4, we discuss how the basic framework can be easily adjusted to accommodate other directions and include sophisticated tools from logic and cognitive science.

2 The setting

We first construct our logical language, building on the following definitions:

Definition 1 (Inference rule). *Given $\phi_1, \ldots, \phi_n, \psi$ in the standard propositional language $\mathcal{L}_P$ (based on a set of atoms Φ), an inference rule R_i is a formula of the form $\{\phi_1, \ldots, \phi_n\} \rightsquigarrow \psi$.*

[1] In [18] an impossible-worlds semantics is presented, but again reasoning is captured via modalities standing for a *number* of steps; this raises concerns analogous to the ones discussed before.

[2] A notable exception where awareness is affected by reasoning is given in [23]; in what follows, we design a rule-based approach but without appealing to a notion of awareness.

Notice that, according to this definition, each R_i stands for an *instance* of a rule (and not for a rule scheme). We then use $pr(R_i)$ and $con(R_i)$ to abbreviate the set of premises and the conclusion of R_i.[3] The rule is to say that whenever the premises are true, the conclusion is also true. We also use $\mathcal{L_R}$ to denote the set of inference rules and $\mathcal{L} := \mathcal{L}_P \cup \mathcal{L_R}$.

Definition 2 (Translation). *The translation of a formula in $\mathcal{L}$ is defined as: $Tr(\phi) := \phi$, if $\phi \in \mathcal{L}_P$ and $Tr(R_i) := \bigwedge_{\phi \in pr(R_i)} \phi \to con(R_i)$, if $R_i \in \mathcal{L_R}$.*

We now define the language of this framework:

Definition 3 (Language $\mathcal{L}_{\mathrm{RB}}$). *Given a countable set of propositional atoms Φ, the language $\mathcal{L}_{\mathrm{RB}}$ is defined inductively as follows:*

$$\phi \quad ::= \quad p \quad | \quad \neg\phi \quad | \quad \phi \wedge \phi \quad | \quad K\psi \quad | \quad \langle R_i \rangle \phi$$

with $p \in \Phi, \psi \in \mathcal{L}, R_i \in \mathcal{L_R}$.

As usual, $K\psi$ reads "the agent knows ψ". $\mathcal{L}_{\mathrm{RB}}$ includes knowledge assertions for *rules* too. That is, apart of knowledge of facts, we can also express which *rules* the agent knows (and is therefore capable of applying). Each $\langle R_i \rangle$ is seen as a labeled operator for a rule-application. A formula $\langle R_i \rangle \phi$ reads "after some application of inference rule R_i, ϕ is true". Dual modalities of the form $[R_i]$ such that $[R_i]\phi$ expresses "after *any* application of R_i, ϕ is true", and the remaining Boolean connectives are defined as usual.

Next, we define our model motivated by the idea that reasoning steps, expressed through rule-applications, should be hardwired in it. We introduce possible worlds that are connected according to the effect of inference rules. Since an agent's reasoning affects the information she holds (rather than truth of facts), the usual valuation function is accompanied by a function yielding which formulas the agent knows at each world. In this sense, each world represents what is explicitly known at it and each rule triggers suitable transitions between them.

Definition 4 (Model). *A model is a tuple $M = \langle W, T, V_1, V_2 \rangle$ where*

- *W is a non-empty set of worlds.*
- *$T : \mathcal{L_R} \to \mathcal{P}(W \times W)$ is a function such that a binary relation on W is assigned to each inference rule in $\mathcal{L_R}$. That is, for $R_i \in \mathcal{L_R}$, $T(R_i) = T_i \subseteq W \times W$, standing for the transition between worlds induced by the rule R_i.*
- *$V_1 : W \to \mathcal{P}(\Phi)$ is a valuation function assigning a set of propositional atoms to each world; intuitively those that are true at the world.*
- *$V_2 : W \to \mathcal{P}(\mathcal{L})$ is a function assigning a set of formulas of $\mathcal{L}$ to each world; intuitively those that the agent knows at the world.*

[3] We emphasize that R_i denotes a *single* rule instance. The rule, which is in fact a pair, composed of the set of premises and the conclusion, is given in terms of the notation $\rightsquigarrow$ for readability and convenience.

The truth clauses are given as follows:

Definition 5 (Truth clauses).

- $M, w \models p$ *if and only if* $p \in V_1(w)$ *for* $p \in \Phi$.
- $M, w \models K\phi$ *if and only if* $\phi \in V_2(w)$.
- $M, w \models \neg\phi$ *if and only if* $M, w \not\models \phi$.
- $M, w \models \phi \wedge \psi$ *if and only if* $M, w \models \phi$ *and* $M, w \models \psi$.
- $M, w \models \langle R_i \rangle \phi$ *if and only if there exists some* $u \in W$ *such that* wT_iu *and* $M, u \models \phi$.

A formula is *valid in a model* if it is true at every world of the model and *valid* if it is valid in the class of all models. However, certain conditions have to be imposed on our initial, general class, to capture the desired effect of rule-applications. To that end, we need the following:

Definition 6 (Propositional truths). *Let* M *be a model and* $w \in W$ *a world of the model. Its set of propositional truths is* $V_1^*(w) = \{\phi \in \mathcal{L}_P \mid M, w \models \phi\}$.

We can now fix an appropriate class of models, denoted by $\mathbf{M}$. For any model M (with $T(R_i) = T_i$ as defined above), $M \in \mathbf{M}$ if and only if:

1. For any inference rule $R_i = \{\phi_1, \ldots, \phi_n\} \rightsquigarrow \psi$, if $w \in W$ is such that $R_i \in V_2(w)$ and $\phi_1, \ldots, \phi_n \in V_2(w)$, then there exists a world $u \in W$ such that wT_iu.
2. For any $w, u \in W$ and inference rule $R_i = \{\phi_1, \ldots, \phi_n\} \rightsquigarrow \psi$, if wT_iu then $R_i \in V_2(w)$, $\phi_1, \ldots, \phi_n \in V_2(w)$ and $V_2(u) = V_2(w) \cup \{\psi\}$.
3. For any $w \in W$ and $\phi \in \mathcal{L}$, if $\phi \in V_2(w)$ then $Tr(\phi) \in V_1^*(w)$.
4. For any $w, u \in W$ and inference rule R_i, if wT_iu then $V_1^*(w) = V_1^*(u)$.

Condition 1 says that if a world represents an epistemic state containing the premises of a known rule R_i, then it must be connected to some other world by the corresponding T_i. Condition 2 says that if w is T_i-connected to u, then it must be that u enriches the epistemic state of w in terms of R_i. This is to ensure that each transition is associated with some addition of a conclusion to an epistemic state. Condition 3 is imposed to guarantee the veridicality of knowledge and the soundness of the known rules.[4] Finally, condition 4 states that T_i-connected worlds are propositionally indiscernible, i.e. transitions stand for purely epistemic actions.

We present some validities that illustrate desirable properties of reasoning processes and will be instrumental for a balanced response against logical omniscience. For notational convenience, we abbreviate sequences of rules as follows:

- $\langle \ddagger \rangle := \langle R_1 \rangle \ldots \langle R_n \rangle$

[4] Recall that $V_2 : W \rightarrow \mathcal{P}(\mathcal{L})$ and that $\mathcal{L} := \mathcal{L}_P \cup \mathcal{L}_{\mathcal{R}}$. Moreover, it should be clear that the world u whose existence is guaranteed by condition 1, is such that it contains the conclusion of R_i, by condition 2, and the rule R_i is necessarily sound due to condition 3.

- $\langle \dagger \rangle := \langle R'_1 \rangle \dots \langle R'_m \rangle$
standing for "after some application of $R_1(R'_1)$, followed by some application of $R_2(R'_2)$, …, followed by some application of $R_n(R'_m)$" (in that order). Similar abbreviations can be defined for the dual cases; for example, by using $[R_1], \dots, [R_n]$ for the first sequence and $[R'_1], \dots, [R'_m]$ for the second.

Theorem 1 (M-validities).

1. *$\langle \ddagger \rangle K\phi \to Tr(\phi)$ is valid in the class **M**. (Factivity)*
2. *$\langle \ddagger \rangle K\phi \to \langle \ddagger \rangle [\dagger] K\phi$ is valid in the class **M**. (Persistence)*
3. *$\langle \ddagger \rangle K\phi \wedge \langle \dagger \rangle K\psi \to \langle \ddagger \rangle \langle \dagger \rangle (K\phi \wedge K\psi)$ is valid in the class **M**. (Merge)*
4. *For any inference rule R_i, $KR_i \wedge \bigwedge_{\phi \in pr(R_i)} K\phi \to \langle R_i \rangle K con(R_i)$ is valid in the class **M**. (Success)*

Proof.
1. Take arbitrary model $M \in \mathbf{M}$ and arbitrary world $w \in W$ of the model. Suppose $M, w \models \langle \ddagger \rangle K\phi$. Unpacking the sequence according to the abbreviation, $M, w \models \langle R_1 \rangle \dots \langle R_n \rangle K\phi$, for the inference rules $R_1, \dots, R_n$. Following Definition 5, there is a world $u_1 \in W$ such that wT_1u_1 and $M, u_1 \models \langle R_2 \rangle \dots \langle R_n \rangle K\phi$. Continuing like that, there is a world $u_n \in W$ such that $u_{n-1}T_nu_n$ and $M, u_n \models K\phi$, which in turn amounts to $\phi \in V_2(u_n)$. Then, by condition 3, $Tr(\phi) \in V_1^*(u_n)$. From condition 4, $Tr(\phi) \in V_1^*(u_{n-1})$. Continuing this process backwards, $Tr(\phi) \in V_1^*(w)$. Therefore $M, w \models Tr(\phi)$. Given the arbitrariness of $M \in \mathbf{M}$ and $w \in W$, we finally conclude that the formula is valid in the class $\mathbf{M}$.
2. Take arbitrary model $M \in \mathbf{M}$ and arbitrary world $w \in W$ of the model. Suppose $M, w \models \langle \ddagger \rangle K\phi$. Unpacking the sequence according to the abbreviation, this amounts to $M, w \models \langle R_1 \rangle \dots \langle R_n \rangle K\phi$. As in the previous case, we obtain a chain $wT_1u_1 \dots u_{n-1}T_nu_n$ such that $M, u_n \models K\phi$, which in turn amounts to $\phi \in V_2(u_n)$ (1). It suffices to show that $M, u_n \models [\dagger]K\phi$, i.e., by repeating the unpacking, now for $[\dagger] = [R'_1] \dots [R'_m]$, that for every world $v_1 \in W$ such that $u_nT'_1v_1$, …, for every world $v_m \in W$ such that $v_{m-1}T'_mv_m$, $M, v_m \models K\phi$, i.e. $\phi \in V_2(v_m)$. Take arbitrary such $v_1, \dots, v_m$. Then due to condition 2 and (1), $\phi \in V_2(v_1)$ and continuing in the same fashion $\phi \in V_2(v_m)$. Therefore, $M, w \models \langle \ddagger \rangle [\dagger] K\phi$, hence $M, w \models \langle \ddagger \rangle K\phi \to \langle \ddagger \rangle [\dagger] K\phi$, as desired.
3. Take arbitrary model $M \in \mathbf{M}$ and arbitrary world $w \in W$ of the model. Suppose $M, w \models \langle \ddagger \rangle K\phi \wedge \langle \dagger \rangle K\psi$. So $M, w \models \langle \ddagger \rangle K\phi$ and $M, w \models \langle \dagger \rangle K\psi$. As above, we obtain a chain $wT_1u_1 \dots u_{n-1}T_nu_n$ such that $M, u_n \models K\phi$, i.e. $\phi \in V_2(u_n)$, and a chain $wT'_1v_1 \dots v_{m-1}T'_nv_m$ such that $M, v_m \models K\psi$, i.e. $\psi \in V_2(v_m)$. The rough idea of the proof is to make use of the conditions of $\mathbf{M}$ to merge the two chains. By condition 2, we know that $V_2(w) \subseteq V_2(u_n)$ and that $V_2(w)$ contains all the premises of rule R'_1, as well as the rule itself. Therefore, $V_2(u_n)$ in turn contains all the premises of rule R'_1 and the rule itself. By conditions 1 and 2, there is a world z_1 such that $u_nT'_1z_1$ and $V_2(z_1) = V_2(u_n) \cup \{con(R'_1)\}$. Now again, by condition 2, $V_2(v_1) =$

$V_2(w) \cup \{con(R'_1)\}$ and since $V_2(w) \subseteq V_2(u_n)$: $V_2(v_1) \subseteq V_2(z_1)$, so we know that z_1 contains the premises for R'_2 and the rule itself. Again by conditions 1 and 2, there is a world z_2 such that $z_1 T'_2 z_2$ and $V_2(z_2) = V_2(z_1) \cup \{con(R'_2)\}$. Continuing like that, the alternations of condition 2 and condition 1, based on the initial assumptions, yield a world z_m such that $z_{m-1} T'_m z_m$ and $V_2(z_m) = V_2(z_{m-1}) \cup \{con(R'_m)\}$ with $V_2(v_m) \subseteq V_2(z_m)$. Therefore $\psi \in V_2(z_m)$. In addition, as the constructed chain is of the form $u_n T'_1 z_1 T'_2 z_2 \dots T'_m z_m$ and due to condition 2, $\phi \in V_2(z_m)$. So $M, z_m \models K\phi \wedge K\psi$, i.e. $M, u_n \models \langle\dagger\rangle(K\phi \wedge K\psi)$. So finally $M, w \models \langle\ddagger\rangle\langle\dagger\rangle(K\phi \wedge K\psi)$, as desired.

4. Take arbitrary model $M \in \mathbf{M}$ and arbitrary world $w \in W$ of the model. Suppose $M, w \models KR_i \wedge \bigwedge_{\phi \in pr(R_i)} K\phi$. Then $R_i \in V_2(w)$ and $\phi \in V_2(w)$, for every $\phi \in pr(R_i)$. Next, from conditions 1 and 2, there is $v \in W$ such that $wT_i v$ and $V_2(v) = V_2(w) \cup \{con(R_i)\}$. As a result, $M, v \models Kcon(R_i)$. Finally, $M, w \models \langle R_i \rangle Kcon(R_i)$, as desired.

Factivity says that whatever comes to be known is true, i.e. only true information or sound rules become known after reasoning, and *Persistence* says that it remains to be known throughout subsequent reasoning processes. *Merge* exemplifies how the agent merges different reasoning processes, thereby coming to know their outcomes. *Success* captures the effect of applying a rule: the conclusion is added in the agent's epistemic stack. As a concrete example, take the validity of $\bigwedge_{R_i = DNE, MP, CI} KR_i \wedge K\neg\neg\phi \wedge K(\phi \to \psi) \to \langle DNE \rangle \langle MP \rangle \langle CI \rangle K(\phi \wedge \psi)$: after successive applications of specific rules, namely *Double Negation Elimination* ($\{\neg\neg\phi\} \rightsquigarrow \phi$), *Modus Ponens* ($\{\phi, \phi \to \psi\} \rightsquigarrow \psi$) and *Conjunction Introduction* ($\{\phi, \psi\} \rightsquigarrow \phi \wedge \psi$), the agent's knowledge is gradually increased.[5]

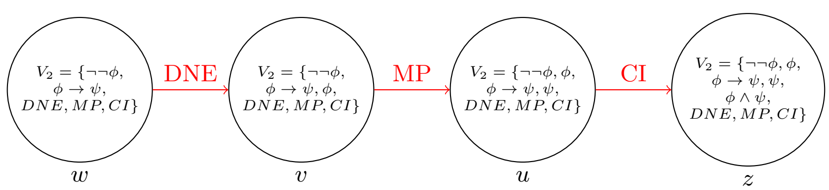

Fig. 1. A model where the reasoning steps of *Double Negation Elimination*, *Modus Ponens*, and *Conjunction Introduction*, taken in this order, correspond to transitions between worlds and reflect how the agent's knowledge is gradually increased.

Logical omniscience is indeed avoided in a balanced way, i.e. still escaping trivialized, totally ignorant agents. The values of knowledge assertions are determined by V_2, which need not obey any closure principle. On the other hand,

[5] We use DNE, MP, and CI to label *particular* instances of Double Negation Elimination, Modus Ponens, and Conjunction Introduction – the ones indicated in parentheses. This labeling only serves the readability of the formulas.

suitable applications of inference rules, reflecting the effort to eventually reach a conclusion, ensure that an agent can *come to know* consequences of her knowledge, provided that she follows the appropriate reasoning track. This is how we avoid an implausible commitment to an automatic and effortless way to expand one's knowledge, as the standard validity $K\phi_1 \wedge \ldots \wedge K\phi_n \rightarrow K\psi$ would dictate. Besides, Cherniak [9] emphasizes that we should view complex deductive reasoning as a task consisting of simple reasoning steps conjoined together. He also argues for a "well-ordering of inferences" in terms of their difficulty, depending both on the rule scheme in question and the logical complexity of its components. Similarly, according to Rips [20], deductive reasoning is a psychological procedure in which sets of formulas are connected via links, that essentially amount to applications of inference rules, just as our framework predicts. Overall, competence is preserved because we unfold the actual processes that result in knowledge and account for their dynamic nature. Logical ignorance is thus ruled out because of a more realistic modelling of the underlying reasoning and not because of ad-hoc restrictions imposed on an inflexible notion of knowledge.

It is interesting to see how our rule-based setting fits in the landscape of similar attempts. As in [1, 14], temporal-style connections encode the progress in the agent's reasoning.[6] Unlike [4, 11, 14, 18], we abstain from a generic notion of reasoning process, instead accounting explicitly for (a) specific rules available to the agent, (b) their individual applications, (c) their chronology, thus monitoring the path that eventually leads to knowledge. This elaborate analysis is, as we remarked above and will further discuss in Section 4, crucial in bridging epistemic frameworks with empirical facts.[7] Furthermore, the enterprise of providing a semantics contributes to Rasmussen's attempt [17], who keeps track of rules applied by the agent, on one hand, but lacks a principled way to assess the validity of his proposed axioms, on the other. Constructing a suitable semantic model that reflects rule-based reasoning gives a concrete view on the credibility of axioms and the adequacy of the solution. Finally, implicit and explicit notions can be discerned, not through an arbitrary filter (as with awareness), but through the analysis of the agent's reasoning.

3 Axiomatization

In this section, we develop the logic Λ_{RB}. We thus obtain a full-fledged logical response against the problem and solid ground to defend our selected axioms.

Definition 7 (Axiomatization of Λ_{RB}). *The axiomatization of Λ_{RB} is given by Table 1.*

Theorem 2 (Soundness). *The logic Λ_{RB} is sound with respect to $\boldsymbol{M}$.*

[6] We note that the frameworks described in [1, 2, 3] that extend the idea of state-transitions to multi-agent settings are particularly interesting for the development of multi-agent variants of our framework too.

[7] More on why this is a worthwhile task can be found in [6].

Table 1.

	AXIOMS
PC	All instances of classical propositional tautologies
K	$[R_i](\phi \to \psi) \to ([R_i]\phi \to [R_i]\psi)$
T	$K\phi \to Tr(\phi)$
Succession	$KR_i \wedge \bigwedge_{\phi \in pr(R_i)} K\phi \to \langle R_i \rangle \top$
Tracking knowledge	$\langle R_i \rangle K\chi \to \bigwedge_{\phi \in pr(R_i)} K\phi \wedge KR_i \wedge K\chi$, for $\chi \neq con(R_i)$
Knowledge of conclusions	$[R_i]Kcon(R_i)$
$Prop_1$	$\langle R_i \rangle \phi \to \phi$, for $\phi \in \mathcal{L}_P$
$Prop_2$	$\phi \to [R_i]\phi$, for $\phi \in \mathcal{L}_P$
Monotonicity	$K\chi \to [R_i]K\chi$
	RULES
Modus Ponens	From ϕ and $\phi \to \psi$, infer ψ
Necessitation	From ϕ infer $[R_i]\phi$

Proof. It suffices to show that the axioms of Definition 7 are valid in the class $\mathbf{M}$, as our rules preserve validity as usual.

- The claim for PC and K is trivial.
- The claim for T follows immediately from condition 3.
- The claim for *Succession* follows from condition 1.
- For *Tracking knowledge*: Take any model $M \in \mathbf{M}$ and world $w \in W$ of the model such that $M, w \models \langle R_i \rangle K\chi$, for $R_i = \{\phi_1, \ldots, \phi_n\} \rightsquigarrow \psi$. So there is $u \in W$ such that wT_iu and $\chi \in V_2(u)$. By condition 2, $\phi_1, \ldots, \phi_n, R_i \in V_2(w)$ and since $V_2(u) = V_2(w) \cup \{\psi\}$, $\chi \in V_2(w) \cup \{\psi\}$. So either $\chi \in V_2(w)$ or $\chi = \psi$. Finally, $M, w \models K\phi_1 \wedge \ldots \wedge K\phi_n \wedge KR_i \wedge K\chi$, for $\chi \neq \psi$.
- The claim for *Knowledge of conclusions* follows from condition 2.
- For $Prop_1$: Take any model $M \in \mathbf{M}$ and world $w \in W$ of the model such that $M, w \models \langle R_i \rangle \phi$ for $\phi \in \mathcal{L}_P$. Then, there is $u \in W$ such that wT_iu and $M, u \models \phi$, i.e. $\phi \in V_1^*(u)$. By condition 4, $\phi \in V_1^*(w)$, i.e. $M, w \models \phi$ as desired.
- For $Prop_2$: Take any model $M \in \mathbf{M}$ and world $w \in W$ of the model such that $M, w \models \phi$. Take any $u \in W$ such that wT_iu. Then by condition 4, $\phi \in V_1^*(u)$, i.e. $M, u \models \phi$ so $M, w \models [R_i]\phi$, as desired.
- For *Monotonicity*: Take any model $M \in \mathbf{M}$ and world $w \in W$ of the model such that $M, w \models K\chi$, i.e. $\chi \in V_2(w)$. Take any $u \in W$ such that wT_iu. From condition 2, $\chi \in V_2(u)$, i.e. $M, u \models K\chi$. But then indeed $M, w \models [R_i]K\chi$.

Aiming at completeness, we follow the procedure of [8], employing *canonical models.*

Lemma 1 (Lindenbaum's Lemma). *If Γ is a Λ_{RB}-consistent set of formulas, then it can be extended to a maximal Λ_{RB}-consistent set Γ^+.*

Proof. The proof goes as usual in these cases. After enumerating $\phi_0, \phi_1, \ldots$, the formulas of our language, one constructs the set Γ^+ as $\bigcup_{n \geq 0} \Gamma^n$ where: $\Gamma^0 = \Gamma$,

$\Gamma^{n+1} = \Gamma^n \cup \{\phi_n\}$, if this is Λ_{RB}-consistent and $\Gamma^n \cup \{\neg\phi_n\}$ otherwise. The desired properties are easily obtained due to this construction.

Definition 8 (Canonical Model). *The canonical model $\mathcal{M}$ for Λ_{RB} is a tuple $\langle \mathcal{W}, \mathcal{T}, \mathcal{V}_1, \mathcal{V}_2 \rangle$ where:*

- $\mathcal{W} = \{w \mid w$ *a maximal Λ_{RB}-consistent set*$\}$.
- $\mathcal{T} : \mathcal{L}_{\mathcal{R}} \to \mathcal{P}(\mathcal{W} \times \mathcal{W})$*, such that for $R_i \in \mathcal{L}_{\mathcal{R}}$, $\mathcal{T}(R_i) = \mathcal{T}_i$, where $w\mathcal{T}_i u$ if and only if* $\{\langle R_i \rangle \phi \mid \phi \in u\} \subseteq w$.
- $\mathcal{V}_1 : \mathcal{W} \to \mathcal{P}(\Phi)$ *such that* $\mathcal{V}_1(w) = \{p \in \Phi \mid p \in w\}$.
- $\mathcal{V}_2 : \mathcal{W} \to \mathcal{P}(\mathcal{L})$ *such that* $\mathcal{V}_2(w) = \{\phi \in \mathcal{L} \mid K\phi \in w\}$.

It is easy to see that an equivalent formulation for the definition of $\mathcal{T}_i$ is $\{\phi \mid [R_i]\phi \in w\} \subseteq u$. Given the definition of the canonical model and our language $\mathcal{L}_{\mathrm{RB}}$, we show:

Lemma 2 (Existence lemma). *For any formula ϕ in our language and $w \in \mathcal{W}$, if $\langle R_i \rangle \phi \in w$ then there is $u \in \mathcal{W}$ such that $w\mathcal{T}_i u$ and $\phi \in u$.*

Proof. Suppose $\langle R_i \rangle \phi \in w$. Take $S = \{\phi\} \cup \{\psi \mid [R_i]\psi \in w\}$. This set is consistent. Were it inconsistent, there would be $\psi_1, \ldots, \psi_n$ such that $\vdash_{\Lambda_{\mathrm{RB}}} \psi_1 \wedge \ldots \wedge \psi_n \to \neg\phi$. Using $[R_i]$-necessitation, distribution and propositional tautologies we obtain $\vdash_{\Lambda_{\mathrm{RB}}} ([R_i]\psi_1 \wedge \ldots \wedge [R_i]\psi_n) \to [R_i]\neg\phi$. By the property of w as maximal consistent set and since $[R_i]\psi_1, \ldots, [R_i]\psi_n \in w$: $[R_i]\neg\phi \in w$. Therefore $\neg\langle R_i \rangle \phi \in w$. Indeed, we have reached a contradiction. Next, we extend S to S^+ according to Lindenbaum's lemma. Then, $\phi \in S^+$ and $[R_i]\psi \in w$ implies $\psi \in S^+$. Take $u := S^+$. As a result, $w\mathcal{T}_i u$ and $\phi \in u$.

Lemma 3 (Truth lemma). *For any formula ϕ in our language and $w \in \mathcal{W}$: $\mathcal{M}, w \models \phi$ if and only if $\phi \in w$.*

Proof. The proof is by induction on the complexity of ϕ.

- Base cases: Consider $\phi := p$ with $p \in \Phi$. Then $\mathcal{M}, w \models p$ if and only if $p \in \mathcal{V}_1(w)$, and by definition, this is the case if and only if $p \in w$. Next, take $\phi := K\psi$ with $\psi \in \mathcal{L}$. Then $\mathcal{M}, w \models K\psi$ if and only if $\psi \in \mathcal{V}_2(w)$, and by definition, this is the case if and only if $K\psi \in w$.
- For $\phi := \neg\psi$ and $\phi := \psi \wedge \chi$, the claim follows easily from I.H. and the maximal consistency of w.
- For $\phi := \langle R_i \rangle \psi$ with I.H. that the result holds for ψ. Then $\mathcal{M}, w \models \langle R_i \rangle \psi$ if and only if there is $u \in \mathcal{W}$ such that $w\mathcal{T}_i u$ and $\mathcal{M}, u \models \psi$. By I.H. this is the case if and only if $\psi \in u$, and by definition of $\mathcal{T}_i$, we get $\langle R_i \rangle \psi \in w$. The other direction follows immediately from the existence lemma.

Theorem 3 (Completeness). *For any set of formulas Γ and formula ϕ in our language: $\Gamma \models_{\boldsymbol{M}} \phi$ only if $\Gamma \vdash_{\Lambda_{\mathrm{RB}}} \phi$.*

Proof.

- We first expand Γ to a maximal Λ_{RB}-consistent set Γ^+. Then, let the canonical model $\mathcal{M}$ be as constructed according to Definition 8. Then by Lemma 3, $\mathcal{M}, \Gamma^+ \models \Gamma$. It suffices to show that $\mathcal{M}$ fulfills the conditions of **M**.
- Condition 1 is satisfied.
  Take inference rule $R_i = \{\phi_1, \ldots, \phi_n\} \rightsquigarrow \psi$ and $w \in \mathcal{W}$ with $R_i, \phi_1, \ldots, \phi_n \in \mathcal{V}_2(w)$, i.e. KR_i, $K\phi_1, \ldots, K\phi_n \in w$ (1). We want to show that there is a world $u \in \mathcal{W}$ such that $w\mathcal{T}_i u$. From (1), $KR_i \wedge K\phi_1 \wedge \ldots \wedge K\phi_n \in w$. But from *Succession*, we get that $\langle R_i \rangle \top \in w$. Using the existence lemma, there is indeed $u \in \mathcal{W}$ such that $w\mathcal{T}_i u$.
- Condition 2 is satisfied.
  Suppose that $w\mathcal{T}_i u$ with $R_i = \{\phi_1, \ldots, \phi_n\} \rightsquigarrow \psi$, i.e. if $\phi \in u$ then $\langle R_i \rangle \phi \in w$. Take arbitrary $\chi \in \mathcal{V}_2(u)$. That is, $K\chi \in u$. Therefore, $\langle R_i \rangle K\chi \in w$. From *Tracking knowledge*, $\phi_1, \ldots, \phi_n, R_i \in \mathcal{V}_2(w)$. From *Knowledge of conclusions* and definition of $\mathcal{T}_i$, $K\psi \in u$, i.e. $\psi \in \mathcal{V}_2(u)$. Furthermore by this definition and *Monotonicity* we obtain that $\mathcal{V}_2(w) \subseteq \mathcal{V}_2(u)$. Therefore, $\mathcal{V}_2(w) \cup \{\psi\} \subseteq \mathcal{V}_2(u)$. Next take $\phi \in \mathcal{V}_2(u)$ with $\phi \neq \psi$. Then $\langle R_i \rangle K\phi \in w$. From *Tracking knowledge*, $K\phi \in w$. As a result, $\phi \in \mathcal{V}_2(w)$. Clearly then, $\mathcal{V}_2(u) = \mathcal{V}_2(w) \cup \{\psi\}$.
- Condition 3 is satisfied.
  Let ϕ be a formula in $\mathcal{L}$. Suppose that $\phi \in \mathcal{V}_2(w)$. That is, $K\phi \in w$. Then by T we obtain, $Tr(\phi) \in w$, that is $\mathcal{M}, w \models Tr(\phi)$ and therefore $Tr(\phi) \in \mathcal{V}_1^*(w)$.
- Condition 4 is satisfied.
  Take $w, u \in \mathcal{W}$ and $w\mathcal{T}_i u$. By definition of $\mathcal{T}_i$, if $\phi \in u$ then $\langle R_i \rangle \phi \in w$. Now take arbitrary $\phi \in \mathcal{L}_P$ such that $\mathcal{M}, u \models \phi$, i.e. $\phi \in \mathcal{V}_1^*(u)$. This means that $\phi \in u$, therefore $\langle R_i \rangle \phi \in w$. From *Prop*$_1$, we obtain $\phi \in w$, i.e. $\mathcal{M}, w \models \phi$ so $\phi \in \mathcal{V}_1^*(w)$. As ϕ was arbitrary, $\mathcal{V}_1^*(u) \subseteq \mathcal{V}_1^*(w)$. For the other inclusion, take arbitrary $\phi \in \mathcal{L}_P$ such that $\mathcal{M}, w \models \phi$, i.e. $\phi \in \mathcal{V}_1^*(w)$. This means that $\phi \in w$. From *Prop*$_2$, we get that $[R_i]\phi \in w$ too. Then we exploit the alternative definition of $\mathcal{T}_i$; since $[R_i]\phi \in w$, $\phi \in u$, i.e. $\mathcal{M}, u \models \phi$ so $\phi \in \mathcal{V}_1^*(u)$. As ϕ was arbitrary, $\mathcal{V}_1^*(w) \subseteq \mathcal{V}_1^*(u)$. Overall, $\mathcal{V}_1^*(w) = \mathcal{V}_1^*(u)$.

4 Extensions

This setting, whose key elements have been hitherto described, can also accommodate more intricate scenarios and facilitate applications informed by other disciplines. In particular, we briefly explain that other tools from (D)EL can be naturally combined with our rule-based logic and that, apart from AI, our syntactic approach can be also relevant for cognitive science.

First, a notion of *implicit* knowledge is not precluded in our framework, for it too employs possible worlds and can be easily endowed with an accessibility relation. Notions of belief can be also included along the lines presented so far, i.e. by simply attaching another function to the model, now yielding the explicit beliefs. Nevertheless, one might drop conditions on factivity or monotonicity. Regarding higher-order knowledge – provided that the language and the range of V_2 are extended – we can also avoid unlimited introspection, as is arguably desired for non-ideal agents. Just as with factual reasoning though, our framework can model *moderate* introspective abilities, via the introduction of introspective rules, whose semantic effect is similarly captured via world transitions.

Moreover, just like *public announcements* of DEL,[8] which may enhance the agent's knowledge, there can be actions for the learning of formulas in $\mathcal{L}$, that is not only of propositional formulas, but also of rules. Their effect is captured by (suitably) tweaking the components of our model to ensure that the formula or rule in question is included. In this way we can bring together external information and the agent's internal reasoning processes. For instance, consider an agent who knows $\phi \to \psi$ and $\neg\psi$, but comes to learn the *Modus Tollens* rule $(\{\phi \to \psi, \neg\psi\} \rightsquigarrow \neg\phi\})$. The combination of this learning action and the application of the rule leads to the agent coming to know that $\neg\phi$ too. Notice that the inclusion of DEL-style operators in this framework still allows for a sound and complete logic. This is because their effect is reducible to formulas not involving such operators. More specifically, *reduction axioms* can gradually "reduce" the truth of complicated formulas in the extended language to the truth of simpler formulas, up until the point where no operator is needed. Provided that these axioms are valid in $\mathbf{M}$, we may simply refer to the completeness of Λ_{RB} and show that a logic built from Definition 7 and the reduction axioms is sound and complete w.r.t. $\mathbf{M}$.

The use of labeled operators and the order-sensitivity of applications of rules make it easier to exploit the observations of cognitive scientists for a precise modelling of resource-bounded reasoners. For example, [15, 20, 22] suggest that not all rules are equally difficult for agents. According to Rips [20], the length and the difficulty of the rules involved in the mental proof constructed for a complex reasoning task determines its overall difficulty. In [19] the need to assign different weights to different rules is experimentally verified and in [24] empirically calculated weights are attached to different rules. Our framework can take these points into consideration. By fixing the agent's capacity (c), attaching empirically indicated weights to rules and introducing inequality formulas to the language (of the form $c \geq c_{R_i}$, where c_{R_i} is intuitively interpreted as the weight of R_i), we can place preconditions to applications of rules and therefore pinpoint where the cutoff of a reasoning process lies.[9]

On a more technical note, while we have presented a Hilbert-style axiomatization of Λ_{RB}, it would be interesting to develop a labeled sequent calculus alternative to this and investigate the proof-theoretic properties of our system. This investigation can be especially relevant to the state-transition settings studying single- or multi-agent reasoning processes. In this way, we can obtain other independent technical results to motivate the use of such systems.

[8] As usual in DEL [5, 10], we can add action operators to our language and capture their effect via model transformations triggered by the action. A formula with dynamic operators, of the form $[\alpha]\phi$, is evaluated by examining what the truth value of ϕ is at a transformed model, obtained via action α.

[9] In fact, this idea can be also pursued along the lines of DEL. The reasoning capacity c of the agent, as an additional component of our models, can be updated (i.e. reduced) following each rule application.

5 Conclusions

We argued that one of the important challenges for epistemic logic is not only to overcome logical omniscience, but to do so while securing the logical competence of agents. We located this endeavour's key parameter in bounded reasoning and spelled it out in logical terms by keeping track of the inference rules the agent applies. We explained how this enriches existing rule-based approaches and expands the scope of their applications. A sound and complete axiomatization was also provided, followed by a summary of our extensions of the core setting.

Acknowledgments. This work is funded by the Dutch Organization for Scientific Research, under the "PhDs in the Humanities" scheme (project number 322-20-018). The author also thanks the audience of the student session of ESSLLI 2018 and the anonymous reviewers for their valuable feedback.

References

[1] Ågotnes, T., Alechina, N.: The dynamics of syntactic knowledge. Journal of Logic and Computation **17**(1), 83–116 (2007)
[2] Ågotnes, T., Walicki, M.: A logic of reasoning, communication and cooperation with syntactic knowledge. In: AAMAS (2005)
[3] Alechina, N., Jago, M., Logan, B.: Modal logics for communicating rule-based agents. In: ECAI (2006)
[4] Alechina, N., Logan, B.: A logic of situated resource-bounded agents. Journal of Logic, Language and Information **18**, 79–95 (2009)
[5] Baltag, A., Renne, B.: Dynamic epistemic logic. In: Zalta, E.N. (ed.) The Stanford Encyclopedia of Philosophy. Metaphysics Research Lab, Stanford University, winter 2016 edn. (2016)
[6] van Benthem, J.: Logic and reasoning: Do the facts matter? Studia Logica: An International Journal for Symbolic Logic **88**(1), 67–84 (2008)
[7] van Benthem, J.: Tell it like it is: Information flow in logic. Journal of Peking University (Humanities and Social Science Edition) **1**, 80–90 (2008)
[8] Blackburn, P., de Rijke, M., Venema, Y.: Modal Logic. Cambridge University Press, New York, NY, USA (2001)
[9] Cherniak, C.: Minimal Rationality. Bradford book, MIT Press (1986)
[10] van Ditmarsch, H., van der Hoek, W., Kooi, B.: Dynamic Epistemic Logic. Springer Publishing Company, Incorporated (2007)
[11] Duc, H.N.: Reasoning about rational, but not logically omniscient, agents. Journal of Logic and Computation **7**(5), 633 (1997)
[12] Fagin, R., Halpern, J.Y.: Belief, awareness, and limited reasoning. Artificial Intelligence **34**(1), 39–76 (1987)
[13] Fagin, R., Halpern, J.Y., Moses, Y., Y., V.M.: Reasoning About Knowledge. MIT press (1995)
[14] Jago, M.: Epistemic logic for rule-based agents. Journal of Logic, Language and Information **18**(1), 131–158 (2009)
[15] Johnson-Laird, P.N., Byrne, R.M., Schaeken, W.: Propositional reasoning by model. Psychological Review **99**(3), 418–439 (1992)

[16] Konolige, K.: A Deduction Model of Belief. Morgan Kaufmann Publishers (1986)
[17] Rasmussen, M.S.: Dynamic epistemic logic and logical omniscience. Logic and Logical Philosophy **24**, 377–399 (2015)
[18] Rasmussen, M.S., Bjerring, J.C.: A dynamic solution to the problem of logical omniscience. Journal of Philosophical Logic (forthcoming)
[19] Rijmen, F., De Boeck, P.: Propositional reasoning: The differential contribution of "rules" to the difficulty of complex reasoning problems. Memory & Cognition **29**(1), 165–175 (2001)
[20] Rips, L.J.: The Psychology of Proof: Deductive Reasoning in Human Thinking. MIT Press, Cambridge, MA, USA (1994)
[21] Stanovich, K.E., West, R.F.: Individual differences in reasoning: Implications for the rationality debate? Behavioral and Brain Sciences **23**(5), 645–665 (2000)
[22] Stenning, K., van Lambalgen, M.: Human Reasoning and Cognitive Science. Boston, USA: MIT Press (2008)
[23] Velázquez-Quesada, F.R.: Small Steps in Dynamics of Information. Ph.D. thesis, Institute for Logic, Language and Computation (ILLC), Amsterdam, The Netherlands (2011)
[24] Zhai, F., Szymanik, J., Titov, I.: Toward probabilistic natural logic for syllogistic reasoning (2015)