Logic and Logical Philosophy
Volume 25 (2016), 499–520
DOI: 10.12775/LLP.2016.016

Ekaterina Kubyshkina
Dmitry V. Zaitsev

RATIONAL AGENCY FROM A TRUTH-FUNCTIONAL PERSPECTIVE

Abstract. The aim of the present paper is to introduce a system, where the epistemic state of an agent is represented truth-functionally. In order to obtain this system, we propose a four-valued logic, that we call the logic of rational agent, where the fact of knowing something is formalized at the level of valuations, without the explicit use of epistemic knowledge operator. On the basis of this semantics, a sound and complete system with two distinct truth-functional negations (an "ontological" and an "epistemic" one) is provided. These negations allow us to express the statements about knowing or not knowing something at the syntactic level. Moreover, such a system is applied to the analysis of knowability paradox. In particular, we show that the paradox is not derivable in terms of the logic of rational agent.

Keywords: many-valued logics; generalized truth values; Church-Fitch's paradox

1. Introduction

It is common to formalize the expressions of the form "agent a knows that p" by the use of an epistemic operator: $\mathsf{K}_a\, p$. Hintikka [13] provides a non-functional semantic interpretation of this operator in terms of possible worlds semantics. His interpretation is intuitively clear when the formalization of the fact of knowing something is represented as syntactic operator K_a. But sometimes the use of this operator may bring to logical anomalies, for example, to the *Knowability Paradox*. We ask then if the use of K_a-operator is the only way to represent the fact

Received November 30, 2015. *Revised* July 13, 2016. *Published online* July 15, 2016

of knowing or not knowing something in formal system. The aim of the present paper is to introduce a system, where the epistemic operator for knowledge (K_a-operator) does not appear, but the fact of knowing or not knowing some truths (or the falsity of some statement) can be defined truth-functionally. There are two principal sources behind this paper.

On the one hand, it contributes to a project of generalized truth values and corresponding logics, more specifically, to its brand new branch labeled as "logics of generalized classical truth values". The underpinning idea and the first appearance of a logic of generalized truth values dates back to the famous "useful four-valued logic" introduced by J. M. Dunn and N. Belnap in the seventies (cf. [7, 3, 4]), though the method was finalized well after by Y. Shramko *et al.* [28]. An underlying generalized valuation system is a result of generalization procedure, followed by a construction of power set of an initial set of truth values introduced with relevant generalization of valuation function to supply a mapping from formulas into generalized values. Over the past decade, such method has grown into powerful philosophical logic tool (cf. Y. Shramko and H. Wansing [29, 30, 31], S. Odintsov and H. Wansing [22]).

Applied to classical truth values, this approach was proposed by Y. Shramko and D. Zaitsev [39] and developed in Zaitsev [40] and S. Wintein and R. A. Muskens [37]. In contrast with Fregean logical tradition truth values are considered to be structured entities each consisting of two truth components: ontological and epistemological ones. Ontological, or referential, truth corresponds to an abstract object denoted by sentence, while epistemological, or inferential, truth represents a property preserved from premises to conclusion in correct reasoning. Logic defined on this semantical basis appeared to be a kind of generalized classical logic with additional expressive power. In particular, its vocabulary contains two specific logical terms for referential and inferential negations. These negations correspond to two operations of semi-Boolean (semi-classical) complementations, where one of them presupposes a changeover only of the referential component without changing the inferential component, while the other one changes inferential truth and leaves the referential truth unaltered. Interestingly, each negation possessing only half of Boolean properties for complementation, whilst in superposition, they give a full-fledged classical (Boolean) complementation.

As typically of negation-stories, this one appeared to be cliff-hanger. Among a number of intriguing issues one concerns the possibility of a logic whose values also consist of two components, but these components are interpreted in a different way. In what follows such a logic with compound epistemically flowered truth values will be presented.

On the other hand, this logic within the manifold of logics of generalized classical values can be deemed as arising from a very different origin. There is a general consensus on it being a rejection of the principle of bivalence that leads to many-valued logic. Without challenging this opinion, it is worth mentioning that there are at least two representations of this principle — a weak and a strong one. The strong formulation of the principle of bivalence states that each sentence is either true or false, that is,

(SPB) *Each sentence takes as its value precisely one of two truth values: truth or falsehood.*

This formulation is indeed strong; in fact it implicitly concomitantly contains three principles. The weak formulation of the principle of bivalence:

(WPB) *There are exactly two possible truth values of a sentence: truth and falsehood.*

The (WPB) characterizes the set of possible values, where the following two principles govern the behavior of valuation function.

The principle of excluded underdetermined values is often confused with the law of excluded middle:

(PEU) *Every sentence takes as its value at least one of two truth values: truth or falsehood.*

And there is a dual principle of excluded overdetermined values, which rejects contradictory assignments:

(PEO) *Every sentence takes as its value at-most-once of two truth values: truth or falsehood.*

This consideration with respect to many-valued logic provides an opportunity subtly delineate any logics on the ground of principles they neglect. For example, stimulated by the philosophical problem of future contingencies, Łukasiewicz intentionally and explicitly rejected the (WPB) in favor of another logical value different from truth and falsehood and complementing two classical values (the value "Possible"). A completely different interpretation of many-valuedness is due to Kleene. He was motivated by his research on partial recursive relations, which sometimes appear to be undefined. To fix this vague situation, he introduces a third value "undefined". In [15] Kleene makes it explicit that the third value should be understood "neither as possible", "nor true and false", "nor neither true, nor false". Rather it should be perceived of as the "absence of information" or "unknown". He does not question the assumption that every proposition is true or false, but assumes that there are propositions whose truth values we do not know at the present moment. In so doing, he saves in a certain sense the (WPB) but rejects the (PEU), the valuation function in his logic is undetermined.

The motivating idea for a logic we present further is to develop Kleene's intuition and consider a valuational system whose values allow capturing into the knowledge state of a rational agent. And it is at that point, when Harry met Sally, the logic under construction is (1) an extension of Kleene's logic presented as (2) a logic of generalized truth values, also known as *Logic of Rational Agent* (LRA).

In Section 2, LRA will be presented semantically within a broader-context generalized values. Section 3 focuses on axiomatization of LRA and a completeness proof for it. Section 4 deals with some useful application of this new epistemic logic: in particular, we propose a formal consideration of the famous *Knowability Paradox* (also known as Church-Fitch's paradox).

2. Semantics for LRA

We construct a many-valued logic that do not reject the weak formulation of bivalence principle, the (WPB), but introduce the distinction between known and unknown truths (or falsities) on the level of valuations. We start our analyze by consideration that there are four possible values for every proposition. These truth values consist of two components: the first component is the value "true" or "false", '**T**' or '**F**', respectively. The second one is a characteristic of the epistemic state of the agent — "known" or "unknown", '**1**' or '**0**', respectively. We multiply our two sets, $\{\mathbf{T}, \mathbf{F}\}$ and $\{\mathbf{1}, \mathbf{0}\}$, and we have a set of four possible values $\mathbf{Q} := \{\mathbf{T1}, \mathbf{T0}, \mathbf{F1}, \mathbf{F0}\}$. The value of a true proposition, that the agent knows, is '**T1**'; of a true proposition of which the truth is unknown by the agent, is '**T0**'; of a false proposition of which the falsity is known by the agent, is '**F1**'; and '**F0**' for a false proposition of which the value is unknown by the agent. The set of designated values would be $\{\mathbf{T1}\}$, the truths that are known to be true.

On the syntactic level we take language $\mathcal{L}$ that consists of the propositional variables p, q, r, s, etc. Now we define logical connectives of the language $\mathcal{L}$. The first 3 connectives are the classical ones: conjunction, disjunction and negation. Let ϕ and ψ be arbitrary formulas. The binary connectives of conjunction and disjunction are defined by the following truth-tables:

$\phi \wedge \psi$	**T1**	**T0**	**F0**	**F1**
T1	**T1**	**T0**	**F0**	**F1**
T0	**T0**	**T0**	**F0**	**F1**
F0	**F0**	**F0**	**F0**	**F1**
F1	**F1**	**F1**	**F1**	**F1**

$\phi \vee \psi$	**T1**	**T0**	**F0**	**F1**
T1	**T1**	**T1**	**T1**	**T1**
T0	**T1**	**T0**	**T0**	**T0**
F0	**T1**	**T0**	**F0**	**F0**
F1	**T1**	**T0**	**F0**	**F1**

The ontological part of the values (**T** or **F**) for conjunctive or disjunctive formulas behaves in the same way that the values **t** and **f** in classical logic. A conjunction is true iff its both conjuncts are true. A disjunction is true iff at least one of its disjuncts is true. What does it happen with the epistemic part of the values? We suppose that the agent knows classical logic. By this we mean that in some cases he can calculate the truth value of a formula even if he does not know the value of both conjuncts (or disjuncts) of the given formula. For example, if we have a conjunctive formula $(\phi \wedge \psi)$, and an agent knows that ϕ is false (this means that the value of the sub-formula ϕ is **F1**), but he does not know the value of the formula ψ, then this agent must know however that the value of the whole formula is **F1** (i.e., it is false and it is known to be false), by the properties of the classical conjunction. We use the same considerations to define disjunction.

The negation operator (we label it the *ontological negation*) is understood also classically. A proposition is false iff its negation is true. This type of negation does not concern the component that describes the epistemic state of an agent (‘**1**’ or ‘**0**’). If a proposition is known to be true (i.e., if it takes the value ‘**T1**’), then its negation should be false, but the agent will still know the falsity of the proposition in consideration (i.e., the value would be ‘**F1**’). The resulting truth-table goes as follows in truth-table:

ϕ	$\neg\phi$
T1	**F1**
T0	**F0**
F0	**T0**
F1	**T1**

The connectives defined by the above truth-tables are not sufficient to express four values on the syntactic level. In order to obtain the stronger system we need to introduce a new connective '$\sim$' (*epistemic negation*). The definition is given in the following truth-table:

ϕ	$\sim\phi$
T1	**T0**
T0	**T1**
F0	**F1**
F1	**F0**

The connective '$\sim$' changes the epistemic state of an agent from known ('**1**') to unknown ('**0**') and vice versa, this is the reason we call it later in the text the *epistemic negation*. The crucial point here is that this operator is not the analogue of "unknown" (i.e., '$\neg\,\mathsf{K}_a\,\phi$' in terms of epistemic logic). We do not give interpretation of this connector '$\sim$' in the natural language, we just apply it to some formula and give the interpretation after the application. For example, suppose that a formula ϕ takes value **T1** (i.e., ϕ is true and the agent knows it), then $\sim\phi$ would take the value **T0** (i.e., ϕ in this case is true, but $\sim\phi$ indicates that the agent does not know that ϕ is true). In this case $\sim\phi$ may be associated with "ϕ and not-known that ϕ", but we can not generalize this interpretation to any formula. Suppose that ϕ takes value **T1**, then $\sim\sim\phi$ takes also the value **T1**. Of course in this case $\sim\sim\phi$ can not be interpreted as "not-known that not-known that ϕ", but as "ϕ is true and the agent knows it". This observation makes clear the difference between $\sim\phi$ in the language of the logic of rational agent and $\neg\,\mathsf{K}_a\,\phi$ in epistemic systems. This underlines the difference between our approach and "classical" formalization of statements containing information about epistemic state of an agent. We do not want to give a new interpretation to modal operators, but we are aiming to find an alternative formalization of statements about knowing or not-knowing something.

So we have constructed a *valuation system* as the matrix $\mathfrak{V}_{\mathrm{LRA}} := \langle \mathbf{Q}, \{\mathbf{T1}\}, \mathcal{C}\rangle$, where $\mathcal{C}$ is the set of functions that are interpretations of propositional connectives defined by the above four truth-tables (of course, $\mathbf{Q}$ is the set of the considered values and $\{\mathbf{T1}\}$ is the set of designated values of this matrix).

We've described some semantical considerations about the logic of rational agent. Now to introduce a logic in a strict sense we need to define the entailment relation. There are many ways to define it, the choice depends on which considerations we take as the fundamental ones. In this paper we want to introduce the entailment relation that captures the following features. First of all, we would like to define a truth-preserving entailment relation. But if we define the entailment relation in a classical way (true premisses should entail true conclusion), we obtain the classical logic, where the distinction between known and unknown truths is absent. So our second consideration is to define the entailment not as "any truth" preserving, but "known truths" preserving. Thus, for the language $\mathcal{L}$ we introduce an entailment relation as the standard consequences relation for the matrix $\mathfrak{V}_{\mathrm{LRA}}$, i.e., for arbitrary formulas ϕ and ψ we put:

$$\phi \models_{\mathfrak{V}_{\mathrm{LRA}}} \psi \overset{\mathrm{df}}{\Longleftrightarrow} \forall h \in \hom(\mathcal{L}, \mathfrak{V}_{\mathrm{LRA}})(h(\phi) = \mathbf{T1} \ \Rightarrow \ h(\psi) = \mathbf{T1}),$$

where $\hom(\mathcal{L}, \mathfrak{V}_{\mathrm{LRA}})$ is the set of all homomorphisms from $\mathcal{L}$ into $\mathfrak{V}_{\mathrm{LRA}}$ such that for any $h \in \hom(\mathcal{L}, \mathfrak{V}_{\mathrm{LRA}})$ and all formulas $\phi, \psi \in \mathcal{L}$ we have:

($\wedge$1) $h(\phi \wedge \psi) = \mathbf{T1}$ iff both $h(\phi) = \mathbf{T1}$ and $h(\psi) = \mathbf{T1}$;
($\wedge$2) $h(\phi \wedge \psi) = \mathbf{T0}$ iff either (i) both $h(\phi) = \mathbf{T0}$ and $h(\psi) = \mathbf{T1}$, or (ii) both $h(\phi) = \mathbf{T1}$ and $h(\psi) = \mathbf{T0}$, or (iii) $h(\psi) = \mathbf{T0} = h(\psi)$;
($\wedge$3) $h(\phi \wedge \psi) = \mathbf{F0}$ iff either (i) both $h(\phi) = \mathbf{F0}$ and $h(\psi) \neq \mathbf{F1}$, or (ii) both $h(\psi) = \mathbf{F0}$ and $h(\phi) \neq \mathbf{F1}$;
($\wedge$4) $h(\phi \wedge \psi) = \mathbf{F1}$ iff either $h(\phi) = \mathbf{F1}$ or $h(\psi) = \mathbf{F1}$.
($\vee$1) $h(\phi \vee \psi) = \mathbf{T1}$ iff either $h(\phi) = \mathbf{T1}$ or $h(\psi) = \mathbf{T1}$;
($\vee$2) $h(\phi \vee \psi) = \mathbf{T0}$ iff either (i) both $h(\phi) \neq \mathbf{T1}$ and $h(\psi) = \mathbf{T0}$, or (ii) both $h(\phi) = \mathbf{T0}$ and $h(\psi) \neq \mathbf{T1}$;
($\vee$3) $h(\phi \vee \psi) = \mathbf{F0}$ iff either (i) $h(\phi) = \mathbf{F0} = h(\psi)$, or (ii) both $h(\phi) = \mathbf{F0}$ and $h(\psi) = \mathbf{F1}$, or (iii) both $h(\phi) = \mathbf{F1}$ and $h(\psi) = \mathbf{F0}$;
($\vee$4) $h(\phi \vee \psi) = \mathbf{F1}$ iff both $h(\phi) = \mathbf{F1}$ and $h(\psi) = \mathbf{F1}$;
($\neg$1) $h(\neg\phi) = \mathbf{T1}$ iff $h(\phi) = \mathbf{F1}$;
($\neg$2) $h(\neg\phi) = \mathbf{T0}$ iff $h(\phi) = \mathbf{F0}$;
($\neg$3) $h(\neg\phi) = \mathbf{F0}$ iff $h(\phi) = \mathbf{T0}$;
($\neg$4) $h(\neg\phi) = \mathbf{F1}$ iff $h(\phi) = \mathbf{T1}$.
($\sim$1) $h(\sim\phi) = \mathbf{T1}$ iff $h(\phi) = \mathbf{T0}$;
($\sim$2) $h(\sim\phi) = \mathbf{T0}$ iff $h(\phi) = \mathbf{T1}$;
($\sim$3) $h(\sim\phi) = \mathbf{F0}$ iff $h(\phi) = \mathbf{F1}$;
($\sim$4) $h(\sim\phi) = \mathbf{F1}$ iff $h(\phi) = \mathbf{F0}$.

The corresponding logical system, the system LRA, would be presented in the next section.

3. System LRA

We determine the logical system that corresponds to the entailment relation $\models_{\mathfrak{V}_{\mathrm{LRA}}}$ and to the model $\mathcal{M}$. For the language $\mathcal{L}$ we introduce some first degree consequence system, which we call LRA. The system is a pair $\langle \mathcal{L}, \vdash_{\mathrm{LRA}} \rangle$, where $\vdash_{\mathrm{LRA}}$ is a binary relation[1] in $\mathcal{L}$ which is the reflexive and transitive closure of the following principles (inference schemes):[2]

$$\phi \wedge \psi \vdash \phi \tag{3.1}$$

$$\phi \wedge \psi \vdash \psi \tag{3.2}$$

$$\phi \vdash \psi \text{ and } \phi \vdash \chi \text{ implies } \phi \vdash \psi \wedge \chi \tag{3.3}$$

$$\phi \vdash \phi \vee \psi \tag{3.4}$$

$$\psi \vdash \phi \vee \psi \tag{3.5}$$

$$\phi \vdash \chi \text{ and } \psi \vdash \chi \text{ implies } \phi \vee \psi \vdash \chi \tag{3.6}$$

$$\phi \wedge (\psi \vee \chi) \vdash (\phi \wedge \psi) \vee (\phi \wedge \chi) \tag{3.7}$$

$$\phi \dashv\vdash \neg\neg\phi \tag{3.8}$$

$$\sim\phi \dashv\vdash \sim\neg\neg\phi \tag{3.9}$$

$$\neg(\phi \wedge \psi) \dashv\vdash \neg\phi \vee \neg\psi \tag{3.10}$$

$$\neg(\phi \vee \psi) \dashv\vdash \neg\phi \wedge \neg\psi \tag{3.11}$$

$$\sim\neg(\phi \vee \psi) \dashv\vdash \sim(\neg\phi \wedge \neg\psi) \tag{3.12}$$

$$\phi \dashv\vdash \sim\sim\phi \tag{3.13}$$

$$\neg\phi \dashv\vdash \neg\sim\sim\phi \tag{3.14}$$

$$\sim\neg\phi \dashv\vdash \neg\sim\phi \tag{3.15}$$

$$\phi \wedge \neg\phi \vdash \psi \tag{3.16}$$

$$\phi \wedge \sim\phi \vdash \psi \tag{3.17}$$

$$\phi \wedge \sim\neg\phi \vdash \psi \tag{3.18}$$

$$\phi \vdash \psi \vee \neg\psi \vee \sim\psi \vee \sim\neg\psi \tag{3.19}$$

$$(\sim\phi \wedge \psi) \vee (\phi \wedge \sim\psi) \vee (\sim\phi \wedge \sim\psi) \dashv\vdash \sim(\phi \wedge \psi) \tag{3.20}$$

$$\begin{aligned}&(\sim\phi \wedge \sim\psi) \vee (\sim\neg\phi \wedge \sim\psi) \vee (\sim\phi \wedge \sim\neg\psi) \vee \\ &\vee (\neg\phi \wedge \sim\psi) \vee (\sim\phi \wedge \neg\psi) \dashv\vdash \sim(\phi \vee \psi)\end{aligned} \tag{3.21}$$

$$\begin{aligned}&(\sim\neg\phi \wedge \psi) \vee (\sim\neg\phi \wedge \sim\psi) \vee (\sim\neg\phi \wedge \sim\neg\psi) \vee \\ &\vee (\sim\phi \wedge \sim\neg\psi) \vee (\phi \wedge \sim\neg\psi) \dashv\vdash \sim\neg(\phi \wedge \psi)\end{aligned} \tag{3.22}$$

---

[1] Here and later we use '$\vdash$' instead of '$\vdash_{\mathrm{LRA}}$' to simplify the reading.

[2] We use '$\phi \dashv\vdash \psi$' as a shorthand for '$\phi \vdash \psi$ and $\psi \vdash \phi$'.

Notice that the postulates that do not contain the epistemic negation '$\sim$' are postulates of Kleene's strong logic $\mathbf{K_3}$. This observation shows that we follow the initial motivation of taking Kleene's intuition about propositions which truth values are unknown.

The system LRA is consistent and complete respect to the entailment relation $\models_{\mathrm{LRA}}$. The proof of consistency of LRA is a routine check of the inference schemes of LRA. We omit this demonstration here.

Theorem 1 (Consistency). *For all $\phi, \psi \in \mathcal{L}$: if $\phi \vdash_{\mathrm{LRA}} \psi$ then $\phi \models_{\mathrm{LRA}} \psi$.*

To prove completeness we need to construct a canonical model. Let an *LRA-theory* be any subset of $\mathcal{L}$ which is closed under $\vdash$ and $\wedge$. More precisely, a subset α of $\mathcal{L}$ is LRA-theory iff for all $\phi, \psi \in \mathcal{L}$:

1. if $\phi \in \alpha$ and $\phi \vdash \psi$, then $\psi \in \alpha$,
2. if $\phi, \psi \in \alpha$, then $\phi \wedge \psi \in \alpha$.

We say that an LRA-theory α is *prime* iff for all $\phi, \psi \in \mathcal{L}$: if $\phi \vee \psi \in \alpha$, then either $\phi \in \alpha$ or $\psi \in \alpha$. Usually, we say that an LRA-theory is *decisive* (or *complete*) iff for each sentence $\phi \in \mathcal{L}$, either $\phi \in \alpha$ or $\neg\phi \in \alpha$. Moreover, similarly we usually say that an LRA-theory α is *consistent* iff there is no $\phi \in \mathcal{L}$ such that $\phi, \neg\phi \in \alpha$.

We define two special properties adopted for the logic of rational agent (LRA). We say that an LRA-theory is *4-decisive* iff for each sentence $\phi \in \mathcal{L}$ either $\phi \in \alpha$, or $\neg\phi \in \alpha$, or $\sim\phi \in \alpha$, or $\sim\neg\phi \in \alpha$. Moreover, we say that a theory α is *4-consistent* iff there is no $\phi \in \mathcal{L}$ such that either $\phi, \neg\phi \in \alpha$, or $\phi, \sim\phi \in \alpha$, or $\phi, \neg\sim\phi \in \alpha$, or $\sim\phi, \neg\phi \in \alpha$. This condition of 4-consistency has the same sense as traditional consistency (restriction of having a sentence with its negation in a theory), but in definition of 4-consistency we take in account that there are two distinct negations. The following lemma due to (3.3) refers to the 4-consistency on the syntactic level:

Proposition 1. *For all $\phi, \psi \in \mathcal{L}$ we have:*

$$\neg\phi \wedge \sim\phi \vdash \psi \tag{$*$}$$

$$\sim\phi \wedge \neg\sim\phi \vdash \psi \tag{$**$}$$

$$\neg\phi \wedge \neg\sim\phi \vdash \psi \tag{$***$}$$

Thus, by the above schemas and (3.16)–(3.18), (3.6), *for all $\phi, \psi \in \mathcal{L}$ we obtain:*

$$(\phi\wedge\neg\phi)\vee(\phi\wedge\sim\phi)\vee(\phi\wedge\neg\sim\phi)\vee(\neg\phi\wedge\sim\phi)\vee(\sim\phi\wedge\neg\sim\phi)\vee(\neg\phi\wedge\neg\sim\phi) \vdash \psi$$

Proof. *Ad* ($*$): First, $\neg\phi \wedge {\sim}\phi \vdash \neg\phi$ and $\neg\phi \vdash {\sim}\neg{\sim}\phi$, by (3.1), (3.14), (3.15), and the transitivity of $\vdash$, respectively. Thus, by the transitivity of $\vdash$, we have $\neg\phi \wedge {\sim}\phi \vdash {\sim}\neg{\sim}\phi$. Second, by (3.2), we have $\neg\phi \wedge {\sim}\phi \vdash {\sim}\phi$. Hence, by (3.3), we obtain $\neg\phi \wedge {\sim}\phi \vdash {\sim}\phi \wedge {\sim}\neg{\sim}\phi$. According to (3.18) we have ${\sim}\phi \wedge {\sim}\neg{\sim}\phi \vdash \psi$. So, applying the transitivity of $\vdash$, we have $\neg\phi \wedge {\sim}\phi \vdash \psi$.

Ad ($**$): it is a substitutional case of the inference scheme (3.16).

Ad ($***$): it follows from $\neg\phi \wedge {\sim}\neg\phi \vdash \psi$ — a substitutional case of the inference scheme (3.17), transitivity of $\vdash$ and schemes (3.1), (3.2), (3.15), and (3.3). $\square$

The following lemma is an analogue of the Lindenbaum's Lemma.

Lemma 1. *Let $\phi \nvdash_{\mathrm{LRA}} \psi$. Then there is a prime 4-decisive and 4-consistent theory α such that $\phi \in \alpha$ and $\psi \notin \alpha$.*

Proof. The existence of a prime LRA-theory α that satisfies the lemma's conditions can be proved in an analogous way as it has been done by Dunn for the system $\mathrm{R_{FDE}}$ in [8], p. 13. We omit this demonstration here. Below we show that α is both 4-decisive and 4-consistent.

First, it is easy to see that α is 4-decisive. Indeed, by (3.19), since α is an LRA-theory and $\phi \in \alpha$, so $\psi \vee \neg\psi \vee {\sim}\psi \vee {\sim}\neg\psi \in \alpha$. So also $\psi \in \alpha$, or $\neg\psi \in \alpha$, or ${\sim}\psi \in \alpha$, or ${\sim}\neg\psi \in \alpha$, since α is also a prime LRA-theory. Thus α is a 4-decisive theory.

To show that α is 4-consistent assume towards contradiction that there is a formula χ such that either $\chi, \neg\chi \in \alpha$, or $\chi, {\sim}\chi \in \alpha$, or $\chi, \neg{\sim}\chi \in \alpha$, or ${\sim}\chi, \neg\chi \in \alpha$. By the fact that α is a prime LRA-theory we have that $(\chi \wedge \neg\chi) \vee (\chi \wedge {\sim}\chi) \vee (\chi \wedge \neg{\sim}\chi) \vee (\neg\chi \wedge {\sim}\chi) \in \alpha$. Hence $\psi \in \alpha$, by (3.3) and Proposition 1, since α is an LRA-theory. Thus α is 4-consistent. $\square$

Now for any prime 4-decisive and 4-consistent LRA-theory α we define the *canonical valuation* c for any propositional variable π as follows:

$$
\begin{aligned}
c(\pi) = \mathbf{T1} &\iff \pi \in \alpha\,,\\
c(\pi) = \mathbf{T0} &\iff {\sim}\pi \in \alpha\,,\\
c(\pi) = \mathbf{F0} &\iff {\sim}\neg\pi \in \alpha\,,\\
c(\pi) = \mathbf{F1} &\iff \neg\pi \in \alpha\,.
\end{aligned}
$$

This valuation c we standardly extend to the homomorphism h^c of $\mathcal{L}$ into the matrix $\mathfrak{V}_{\mathrm{LRA}}$. Notice that we obtain:

Lemma 2. *For any* $\phi \in \mathcal{L}$:

$$h^c(\phi) = \mathbf{T1} \iff \phi \in \alpha\,,$$
$$h^c(\phi) = \mathbf{T0} \iff {\sim}\phi \in \alpha\,,$$
$$h^c(\phi) = \mathbf{F0} \iff {\sim}\neg\phi \in \alpha\,,$$
$$v^c(\phi) = \mathbf{F1} \iff \neg\phi \in \alpha\,.$$

Proof. We use induction on the construction of formulas.

$h^c(\phi \wedge \psi) = \mathbf{T1}$ iff $h^c(\phi) = \mathbf{T1}$ and $h^c(\psi) = \mathbf{T1}$ iff $\phi \wedge \psi \in \alpha$ iff $\phi \wedge \psi \in \alpha$; since α is an LRA-theory and, respectively, by ($\wedge$1), inductive hypothesis, (3.1), and (3.2).

$h^c(\phi \wedge \psi) = \mathbf{T0}$ iff either both $h^c(\phi) = \mathbf{T0}$ and $h^c(\psi) = \mathbf{T1}$, or both $h^c(\phi) = \mathbf{T1}$ and $h^c(\psi) = \mathbf{T0}$, or $h^c(\psi) = \mathbf{T0} = h^c(\psi)$ iff either ${\sim}\phi, \psi \in \alpha$, or $\phi, {\sim}\psi \in \alpha$, or ${\sim}\phi, {\sim}\psi \in \alpha$ iff either ${\sim}\phi \wedge \psi \in \alpha$, or $\phi \wedge {\sim}\psi \in \alpha$, or ${\sim}\phi \wedge {\sim}\psi \in \alpha$ iff $({\sim}\phi \wedge \psi) \vee (\phi \wedge {\sim}\psi) \vee ({\sim}\phi \wedge {\sim}\psi) \in \alpha$ iff ${\sim}(\phi \wedge \psi) \in \alpha$; since α is a prime LRA-theory and, respectively, by ($\wedge$2), inductive hypothesis, (3.1), (3.2), (3.20), and (3.4).

$h^c(\phi \wedge \psi) = \mathbf{F0}$ iff either both $h^c(\phi) = \mathbf{F0}$ and $h^c(\psi) \neq \mathbf{F1}$, or both $h^c(\psi) = \mathbf{F0}$ and $h^c(\phi) \neq \mathbf{F1}$ iff either (both $h^c(\phi) = \mathbf{F0}$ and either $h^c(\psi) = \mathbf{T1}$, or $h^c(\psi) = \mathbf{T0}$, $h^c(\psi) = \mathbf{F0}$), or (both $h^c(\psi) = \mathbf{F0}$ and either $h^c(\phi) = \mathbf{T1}$, or $h^c(\phi) = \mathbf{T0}$, or $h^c(\phi) = \mathbf{F0}$) iff either both $h^c(\phi) = \mathbf{F0}$ and $h^c(\psi) = \mathbf{T1}$, or both $h^c(\phi) = \mathbf{F0}$ and $h^c(\psi) = \mathbf{T0}$, or $h^c(\phi) = \mathbf{F0} = h^c(\psi)$, or both $h^c(\psi) = \mathbf{F0}$ and $h^c(\phi) = \mathbf{T1}$, or both $h^c(\psi) = \mathbf{F0}$ and $h^c(\phi) = \mathbf{T0}$ iff either ${\sim}\neg\phi, \psi \in \alpha$, or ${\sim}\neg\phi, {\sim}\psi \in \alpha$, or ${\sim}\neg\phi, {\sim}\neg\psi \in \alpha$, or ${\sim}\neg\psi, \phi \in \alpha$, or ${\sim}\neg\psi, {\sim}\phi \in \alpha$ iff either ${\sim}\neg\phi \wedge \psi \in \alpha$, or ${\sim}\neg\phi \wedge {\sim}\psi \in \alpha$, or ${\sim}\neg\phi \wedge {\sim}\neg\psi \in \alpha$, or ${\sim}\neg\psi \wedge \phi \in \alpha$, or ${\sim}\neg\psi \wedge {\sim}\phi \in \alpha$ iff $({\sim}\neg\phi \wedge \psi) \vee ({\sim}\neg\phi \wedge {\sim}\psi) \vee ({\sim}\neg\phi \wedge {\sim}\neg\psi) \vee ({\sim}\phi \wedge {\sim}\neg\psi) \vee (\phi \wedge {\sim}\neg\psi) \in \alpha$ iff ${\sim}\neg(\phi \wedge \psi) \in \alpha$; since α is a prime LRA-theory and, respectively, by ($\wedge$3), inductive hypothesis, (3.1), (3.2), (3.4), and (3.22).

$h^c(\phi \wedge \psi) = \mathbf{F1}$ iff either $h^c(\phi) = \mathbf{F1}$ or $h^c(\psi) = \mathbf{F1}$ iff either $\neg\phi \in \alpha$ or $\neg\psi \in \alpha$ iff $\neg\phi \vee \neg\psi \in \alpha$ iff $\neg(\phi \wedge \psi) \in \alpha$; , since α is a prime LRA-theory and, respectively, by ($\wedge$4), inductive hypothesis, (3.4), and (3.10).

$h^c(\phi \vee \psi) = \mathbf{T1}$ iff either $h^c(\phi) = \mathbf{T1}$ or $h^c(\psi) = \mathbf{T1}$ iff either $\phi \in \alpha$ or $\psi \in \alpha$ iff $\phi \vee \psi \in \alpha$; since α is a prime LRA-theory and, respectively, by ($\vee$1), inductive hypothesis, and (3.4).

$h^c(\phi \vee \psi) = \mathbf{T0}$ iff either both $h^c(\phi) \neq \mathbf{T1}$ and $h^c(\psi) = \mathbf{T0}$, or both $h^c(\phi) = \mathbf{T0}$ and $h^c(\psi) \neq \mathbf{T1}$ iff either (both $h^c(\psi) = \mathbf{T0}$ and either $h^c(\phi) = \mathbf{T0}$, or $h^c(\phi) = \mathbf{F0}$, or $h^c(\phi) = \mathbf{T1}$), or (both $h^c(\phi) = \mathbf{T0}$ and either $h^c(\psi) = \mathbf{T0}$, or $h^c(\psi) = \mathbf{F0}$, or $h^c(\psi) = \mathbf{F1}$) iff either ${\sim}\phi, {\sim}\psi \in \alpha$, or $\sim\neg\phi, \sim\psi \in \alpha$, or $\neg\phi, \sim\psi \in \alpha$, or $\sim\phi, \sim\psi \in \alpha$, or $\sim\phi, \sim\neg\psi \in \alpha$, or $\sim\phi, \neg\psi \in \alpha$ iff either $\sim\phi \wedge \sim\psi \in \alpha$, or $\sim\neg\phi \wedge \sim\psi \in \alpha$, or $\neg\phi \wedge \sim\psi \in \alpha$, or $\sim\phi \wedge \sim\neg\psi \in \alpha$, or $\sim\phi \wedge \neg\psi \in \alpha$ iff $(\sim\phi \wedge \sim\psi) \vee (\sim\neg\phi \wedge \sim\psi) \vee (\neg\phi \wedge \sim\psi) \vee (\sim\phi \wedge \sim\neg\psi) \vee (\sim\phi \wedge \neg\psi) \in \alpha$ iff $\sim(\phi \vee \psi) \in \alpha$; since α is a prime LRA-theory and, respectively, by ($\vee$2), inductive hypothesis, (3.1), (3.2), (3.4), and (3.21).

$h^c(\phi\vee\psi) = \mathbf{F0}$ iff either $h^c(\phi) = \mathbf{F0} = h^c(\psi)$, or both $h^c(\phi) = \mathbf{F0}$ and $h^c(\psi) = \mathbf{F1}$, or both $h^c(\phi) = \mathbf{F1}$ and $h^c(\psi) = \mathbf{F0}$ iff either $\sim\neg\phi, \sim\neg\psi \in \alpha$, or $\sim\neg\phi, \neg\psi \in \alpha$, or $\neg\phi, \sim\neg\psi \in \alpha$ iff either $\sim\neg\phi \wedge \sim\neg\psi \in \alpha$, or $\sim\neg\phi \wedge \neg\psi \in \alpha$, or $\neg\phi \wedge \sim\neg\psi \in \alpha$ iff $(\sim\neg\phi\wedge\neg\psi)\vee(\neg\phi\wedge\sim\neg\psi)\vee(\sim\neg\phi\wedge\sim\neg\psi) \in \alpha$ iff $\sim(\neg\phi\wedge\neg\psi) \in \alpha$ iff $\sim\neg(\phi\vee\psi) \in \alpha$; since α is a prime LRA-theory and, respectively, by ($\vee$3), inductive hypothesis, (3.1), (3.2), (3.4), (3.20), and (3.12).

$h^c(\phi \vee \psi) = \mathbf{F1}$ iff $h^c(\phi) = \mathbf{F1} = h^c(\psi)$ iff $\neg\phi, \neg\psi \in \alpha$ iff $\neg\phi \wedge \neg\psi \in \alpha$ iff $\neg(\phi \vee \psi) \in \alpha$; since α is an LRA-theory and, respectively, by ($\vee$4), inductive hypothesis, (3.1), (3.2), and (3.11).

$h^c(\neg\phi) = \mathbf{T1}$ iff $h^c(\phi) = \mathbf{F1}$ iff $\neg\phi \in \alpha$; respectively by ($\neg$1) and inductive hypothesis.

$h^c(\neg\phi) = \mathbf{T0}$ iff $h^c(\phi) = \mathbf{F0}$ iff $\sim\neg\phi \in \alpha$; respectively by ($\neg$2) and inductive hypothesis.

$h^c(\neg\phi) = \mathbf{F0}$ iff $h^c(\phi) = \mathbf{T0}$ iff $\sim\phi \in \alpha$ iff $\neg\sim\neg\phi \in \alpha$ iff $\sim\neg\neg\phi \in \alpha$; since α is an LRA-theory and, respectively, by ($\neg$3), inductive hypothesis, (3.9), and (3.15).

$h^c(\neg\phi) = \mathbf{F1}$ iff $h^c(\phi) = \mathbf{T1}$ iff $\phi \in \alpha$ iff $\neg\neg\phi \in \alpha$; since α is an LRA-theory and, respectively, by ($\neg$4), inductive hypothesis, and (3.8).

$h^c(\sim\phi) = \mathbf{T1}$ iff $h^c(\phi) = \mathbf{T0}$ iff $\sim\phi \in \alpha$; since α is an LRA-theory and, respectively, by ($\sim$1) and inductive hypothesis.

$h^c(\sim\phi) = \mathbf{T0}$ iff $h^c(\phi) = \mathbf{T1}$ iff $\phi \in \alpha$ iff $\sim\sim\phi \in \alpha$; since α is an LRA-theory and, respectively, by ($\sim$2), inductive hypothesis, and (3.13).

$h^c(\sim\phi) = \mathbf{F0}$ iff $h^c(\phi) = \mathbf{F1}$ iff $\neg\phi \in \alpha$ iff $\sim\neg\sim\phi \in \alpha$; since α is an LRA-theory and by ($\sim$3), inductive hypothesis, (3.14), (3.15), respectively.

$h^c(\sim\phi) = \mathbf{F1}$ iff $h^c(\phi) = \mathbf{F0}$ iff $\sim\neg\phi \in \alpha$ iff $\neg\sim\phi \in \alpha$; since α is an LRA-theory and by ($\sim$4), inductive hypothesis, (3.15), respectively. $\square$

Now we can prove the completeness of the system RA.

Theorem 2 (Completeness). *For all $\phi, \psi \in \mathcal{L}$:*

if $\phi \models_{\mathrm{LRA}} \psi$ then $\phi \vdash_{\mathrm{LRA}} \psi$.

Proof. By contraposition. Assume that $\phi \nvdash_{\mathrm{LRA}} \psi$. Then, by Lemma 1, there exists a 4-decisive, 4-consistent prime theory α that both $\phi \in \alpha$ and $\psi \notin \alpha$. Now, by Lemma 2, for the canonical valuation c in this theory we have $h^c(\phi) = \mathbf{T1}$ and $h^c(\psi) \neq \mathbf{T1}$. So $\phi \not\models_{\mathrm{LRA}} \psi$. □

4. Knowability paradox

The system LRA can serve as a truth-functional basis for a new approach to formalization of the statements about an agent's knowledge and ignorance. We propose to apply this promising tool to a widely discussed epistemological issue, known as the Knowability Paradox.

The Knowability Paradox (also known as Fitch's paradox or Church-Fitch's paradox) is a logical argument suggesting that if every true proposition is knowable (the knowability principle), then all true propositions are already known (the omniscience principle).

More precisely, the proof of the paradox is based on the assumption that there is a truth that is not known yet (so called, unknown truth). In terms of epistemic logic we can formalize this assumption as follows[3]:

$$p \wedge \neg \mathsf{K}_a\, p, \tag{$\star$}$$

where p is an arbitrary true proposition, K_a is an epistemic operator that applied to a sentence p denotes "the proposition p is known (to be true) by an agent a". Thus the formula ($\star$) can be interpreted as: "p is the case, but the agent doesn't know it".

Next goes the principle that every true statement is capable of being known to be true (*the knowability principle*):

$$p \rightarrow \Diamond \mathsf{K}_a\, p, \tag{KP}$$

where '$\Diamond$' is a modal operator of possibility. Now consider the application of the above principle to ($\star$), which leads to:

$$\Diamond \mathsf{K}_a(p \wedge \neg \mathsf{K}_a\, p).$$

So, by distribution of the operator 'K_a' over conjunction, one gets:

$$\Diamond(\mathsf{K}_a\, p \wedge \mathsf{K}_a \neg \mathsf{K}_a\, p).$$

[3] We provide here only a sketch of the proof in propositional language without quantifying, as the proof in terms of epistemic logic with appellation to possible-worlds semantics does not make the subject of the current paper.

The formula '$\mathsf{K}_a \neg \mathsf{K}_a\, p$' can be reduced to '$\neg \mathsf{K}_a\, p$' according to "factivity principle", which states that if we know some fact, it takes place (i.e., '$\mathsf{K}_a\, q \to q$'). Thus we arrive at:

$$\Diamond(\mathsf{K}_a\, p \wedge \neg \mathsf{K}_a\, p).$$

The last formula is a substitutional case of '$\Diamond \bot$', that contradicts the commonly accepted principle of the impossibility of contradiction (i.e., '$\neg\Diamond\bot$'). Summing up, our assumption ($\star$) appeared to lead to contradiction and must be discharged:

$$\neg(p \wedge \neg \mathsf{K}_a\, p).$$

Applying PC we can transform the last formula into:

$$p \to \mathsf{K}_a\, p, \tag{OP}$$

where the former means that if the proposition p is true, then its truth is already known (*the omniscience principle*).

Quite predictably, there is a huge number of possible solutions to the paradox. Schematically we can divide the solution strategies into three categories (we follow here the classification given by Maffezioli, Naibo and Negri in [19]):

1. There is a suggestion to restrict the knowability principle. This approach is developed by Dummet [9], Tennant [32], [33], Restall [26].
2. One can also try to reformulate the knowability principle (see e.g. Edgington [11], Martin-Löf [21], Burgess [5], Proietti and Sandu [25]).
3. One may revise the logical framework to avoid the paradox (see e.g. Williamson [35], Beall [1], [2], Wansing [34], Dummet [10], Priest [24]).

In what follows we reformulate the knowability principle in terms of the logic of rational agent. So partially we follow the second strategy, but in so doing we also revise the logical system, as it was proposed by the representatives of the third approach.

For a start, let us list the epistemological and logical principles used in the paradox formulation. Among those, the knowability principle, the principle of distribution of knowledge over conjunction, the factivity principle and the impossibility of contradiction. Our first aim is thus to show that accepting the knowability principle does not lead to the omniscience principle. The second aim of this section is to analyze the other principles involved in the "classical" derivation of the paradox in terms of LRA.

As noted above, the knowability principle has been formalized in epistemic logic as (KP): $p \to \Diamond \mathsf{K}_a\, p$. What does it mean? Literally, "if the proposition is true, then we have a possibility to know it". In the logic of rational agent it can be reformulated as "if the truth value of the proposition was '**T1**' or '**T0**' (the proposition is true without connection to the knowledge the agent may have or not about proposition in consideration), then it is possible for this proposition to be true and known to be true (to take value '**T1**')".

Having two negations (the "ontological" one and the "epistemic" one) at hand, in the system LRA, we have a tool to formalize the fact of knowing or not knowing something by an agent on the syntactic level. To formalize the possibility of knowing something we should introduce modal operators of possibility and necessity with corresponding possible worlds semantics. The resulting language is denoted by $\mathcal{L}_{\Diamond\Box}$. We fix a set of possible worlds W and a binary accessibility relation R in W, which is a reflexive, transitive and symmetric. Moreover, we use *relational models* of the form $\langle W, R, V\rangle$, where V is a valuation function from $\mathcal{L}_{\Diamond\Box} \times W$ into $\{\mathbf{T1}, \mathbf{T0}, \mathbf{F0}, \mathbf{F1}\}$ such that for '$\neg$', '$\sim$', '$\wedge$', and '$\vee$' the valuation V satisfies the same inductive conditions for homomorphisms given on the page 505 and moreover:

($\Box$1) $V(\Box\phi, w) = \mathbf{T1}$ iff $\forall u(wRu \Rightarrow V(\phi, u) = \mathbf{T1})$;
($\Box$2) $V(\Box\phi, w) = \mathbf{T0}$ iff $\forall u(wRu \Rightarrow (V(\phi, u) = \mathbf{T1}$ or $V(\phi, u) = \mathbf{T0}))$ and $\exists u(wRu$ and $V(\phi, u) = \mathbf{T0})$;
($\Box$3) $V(\Box\phi, w) = \mathbf{F1}$ iff $\exists u(wRu$ and $V(\phi, u) = \mathbf{F1})$;
($\Box$4) $V(\Box\phi, w) = \mathbf{F0}$ iff $\forall u(wRu \Rightarrow V(\phi, u) \neq \mathbf{F1})$ and $\exists u(wRu$ and $V(\phi, u) = \mathbf{F0})$.
($\Diamond$1) $V(\Diamond\phi, w) = \mathbf{T1}$ iff $\exists u(wRu$ and $V(\phi, u) = \mathbf{T1})$;
($\Diamond$2) $V(\Diamond\phi, w) = \mathbf{T0}$ iff $\forall u(wRu \Rightarrow V(\phi, u) \neq \mathbf{T1})$ and $\exists u(wRu$ and $V(\phi, u) = \mathbf{T0})$;
($\Diamond$3) $V(\Diamond\phi, w) = \mathbf{F1}$ iff $\forall u(wRu \Rightarrow V(\phi, u) = \mathbf{F1})$;
($\Diamond$4) $V(\Diamond\phi, w) = \mathbf{F0} \Leftrightarrow \forall u(wRu \Rightarrow (V(\phi, u) = \mathbf{F1}$ or $V(\phi, u) = \mathbf{F0}))$ and $\exists u(wRu$ and $V(\phi, u) = \mathbf{F0})$.

Of course, we have $V(\Diamond\phi, w) = V(\neg\Box\neg\phi, w)$. Thus, '$\Diamond$' could be defined by means of '$\Box$'.

By introducing interpretations of modal operators we limit the way of the informal reading of four values of the logic of rational agent. More precisely, the values read as "true and known to be true" (**T1**) and others refer only to formulas having non-modal operators as the main functor. For the modal operators, the reading of the values is defined as above. For example, $V(\Diamond p, w) = \mathbf{T1}$ does not mean "the possibility of p is true and known to be true in the world w", but it means that there exists an accessible from w world, where p is true and known to be true, or "it is possible that p is true and known".

For an axiomatization to be full-fledged, one definitely needs to add to LRA the inference schemes corresponding to appropriate deductive postulates. However, it is a topic in its own right, and we will not address it here. In what follows we use the semantical entailment $\models_{\mathrm{LRA}_{\Diamond\Box}}$ that always leads from '**T1**' to '**T1**' and is an extension of $\models_{\mathrm{LRA}}$.

As noted above, the knowability principle states that if some proposition is true, then it is possible that this proposition would be known to be true. In terms of LRA semantics, this means that if the value of some proposition ϕ is '**T1**' or '**T0**' in the (actual) world w, then there exists an accessible from w world u, where $V(\phi, u) = \mathbf{T1}$ (that is the condition for $V(\Diamond\phi, w) = \mathbf{T1}$). Taking in account interpretations, the former statement (KP) can be put as follows:

$$\phi \vee {\sim}\phi \models_{\mathrm{LRA}_{\Diamond\Box}} \Diamond\phi, \tag{KP$'$}$$

If the left part of the inference takes value '**T1**' (it is the case only if ϕ takes value '**T1**' or '**T0**'), then $\Diamond\phi$ takes value "**T1**" (there exists an accessible world, where ϕ is true and known to be true). With this formulation of the knowability principle in mind, we can reconstruct the derivation of the paradox, now in terms of the logic of rational agent. We start with the same assumption, that there is a proposition that is true and unknown by an agent: ${\sim}p$. If ${\sim}p$ is true and known, then p is true, independently of the agent's state of knowledge:

$${\sim}p \models_{\mathrm{LRA}_{\Diamond\Box}} p \vee {\sim}p.$$

After we apply the knowability principle (KP$'$):

$$p \vee {\sim}p \models_{\mathrm{LRA}_{\Diamond\Box}} \Diamond p$$

and by this we have:

$${\sim}p \models_{\mathrm{LRA}_{\Diamond\Box}} \Diamond p \tag{$\dagger$}$$

that means that if p is true and unknown, then there exists an accessible world, where p is true and known to be true (takes value '**T1**').

Possible objection can raise on the ground that even if application of the knowability principle to the assumption that there is an unknown truth does not lead us to the omniscience principle, it does not mean that the omniscience principle is not derivable in LRA. Fortunately, it is not the case. First, we consider omniscience principle as it can be presented by means of LRA, and then we show it to be underivable in LRA.

The omniscience principle states that all truths are known. In other words, if some proposition is true ('$\phi \vee \sim\phi$' takes value '**T1**'), then it is known to be true (then ϕ takes value '**T1**'):

$$\phi \vee \sim\phi \vdash \phi\,. \tag{OP$'$}$$

However, it is obvious that this inference is not LRA-valid (let ϕ to take the value '**T0**', then the premise takes value '**T1**', while conclusion takes value '**T0**'). Nevertheless, it may still be possible that (KP) together with modal extension of LRA can validate it. As we remarked before the enrichment of LRA by modal operators is left for the future work. In what follows we consider that such an extension should be conservative with respect to the present semantics. Thus, if one considers only conservative extensions of LRA, the principle (OP') will remain not LRA-valid.

As our considerations of knowability and omniscience in terms of LRA has shown, some important logical and epistemological principles involved in standard form of the paradox remain untouched upon. Meanwhile, these principles are often considered as basic and important for philosophy and epistemology. It gives rise to a reasonable question, whether these principles are valid in the logic we propose. The answer is positive, in as much as they can be presented by means of the logic of rational agent. Let's take a closer look at them.

The principle of distributivity of knowledge over conjunction states that if an agent knows that both ϕ and ψ, then he knows that ϕ and he knows that ψ. The acceptability of this principle is still under intense discussion (for detail, consult e.g. Nozick [20], Williamson [36], Duc [6]). The best way to avoid such a discussion is to note that this principle is not only admissible in LRA, but corresponding inference schemes are exactly deductive postulates (3.1) and (3.2) of LRA. These inferences mean that if $\phi \wedge \psi$ is true and known to be true, then ϕ is true and known to be true, and ψ is true and known to be true.

The factivity principle, as avowed above, states that if something is known to be true, then it is true. The same idea may be expressed in

LRA as follows (see (3.4)):

$$\phi \vdash_{\text{LRA}} \phi \vee {\sim}\phi$$

If ϕ is true and known to be true (takes value '**T1**'), then it is true independently of the state of the agent's knowledge (i.e., $\phi \vee {\sim}\phi$ takes value '**T1**'). Thus, the factivity principle in such formulation is also a core principle of LRA.

The last principle to analyze is the impossibility of contradiction ($\neg\Diamond\bot$). To clarify the status of this principle in LRA we should define the contradiction ($\bot$) in the system LRA and its modal extension. The definition of contradiction (or logical falsity) in many-valued logics is an open question. S. Gottwald [12] explicits two possibilities of generalizing the notion of contradiction from classical logic to many-valued logic: one can take as logical falsities all those formulas ϕ for which $\neg\phi$ is a logical truth (takes the designated truth-value); or one can take as logical falsities all those formulas ϕ which assume only antidesignated truth degrees. For the reason of simplicity, we prefer to use the first method to define the contradiction. Having the set of designated values $\{\mathbf{T1}\}$, logical falsities may be defined as all formulas ϕ for which $h(\neg\phi) = \mathbf{T1}$, that gives us $h(\phi) = \mathbf{F1}$. Following this way, one may define $\bot$ as a formula that takes the truth-value '**F1**' in all possible worlds for the modal extension of LRA. Having this definition, it is obvious that for an arbitrary formula ϕ, $\phi \models_{\text{LRA}_{\Diamond\Box}} \neg\Diamond\bot$, that means that such an extension validates the principle of the impossibility of contradiction.

In this section we have shown that we can not derive the omniscience principle from the knowability principle in some conservative modal extension of LRA and that the system contains the other important philosophical and epistemological principles, that are used in the derivation of the knowability paradox. Summing up, in LRA, the knowability principle does not lead to the omniscience, while other important philosophical and epistemological principles involved in knowability paradox are preserved.

5. Conclusion

In this paper, we have introduced a generalization of classical truth values applied to Kleene's idea of undefined values. This generalization gives rise to a four-valued logic, that we label as "logic of rational agent"

(LRA). In so doing, we presented a sound and complete formalization the system LRA of our semantical considerations.

It appeared that the system LRA may be seen as a sort of epistemic logic, taking in account the fact that it provides a formalization of the statements about knowledge and ignorance. We do not insist that the logic we propose must replace "classical" epistemic logic. However, we consider our approach as a possible step towards the clarification of important epistemological and philosophical problems related to the notions of knowledge and ignorance. We have shown that the logic of rational agent as presented above is free from Knowability Paradox.

The future work concerns, first, a choice of intuitively suitable axioms for modal version of LRA and proof of its semantical adequacy; and secondly, further exploring the other possible formalizations of the logic of rational agent, which may be useful in philosophical analyses of the notions of knowledge and ignorance.

Acknowledgments. We would like to thank an anonymous referee as well as Editors of *Logic and Logical Philosophy* journal for helpful comments and suggestions of improvements of the earlier version of this paper. In particular, we are grateful for indicating the sudden gaps in the proofs of Proposition 1 and Lemma 2, and instructive suggestions how to bridge these gaps. Without their invaluable help and friendly assistance this paper would not be properly presented.

References

[1] Beall, J. C., "Fitch's proof, verificationism, and the knower paradox", *Australasian Journal of Philosophy*, 78 (2000): 241–247. DOI: 10.1080/00048400012349521

[2] Beall, J. C., "Knowability and possible epistemic oddities", pages 105–125, in [27], 2009. DOI: 10.1093/acprof:oso/9780199285495.003.0009

[3] Belnap, N. D., "How a computer should think", pages 30–55 in *Contemporary aspects of philosophy*, G. Ryle (ed.), Oriel Press Ltd, Stocksfield, 1977.

[4] Belnap, N. D., "A useful four-valued logic", pages 5–37 in *Modern Uses of Multiple-valued Logic*, M. Dunn and G. Epstein (eds.), volume 2 of the series "Episteme", D. Reidel Publishing Company, Dordrecht, 1977. DOI: 10.1007/978-94-010-1161-7_2

[5] Burgess, J., 2009, "Can truth out?", pages 147–162 in [27], 2009. DOI: 10.1017/CBO9780511487347.012

[6] Duc, H. N., "Reasoning about rational, but not logically omniscient, agents", *Journal of Logic and Computation*, 7, 5 (1997): 633–648. DOI: 10.1093/logcom/7.5.633

[7] Dunn, J. M., "Intuitive semantics for first-degree entailments and 'coupled trees' ", *Philosophical Studies*, 29 (1976): 149–168. DOI: 10.1007/BF00373152

[8] Dunn, J. M., "Partiality and its dual", *Studia Logica*, 66 (2000): 5–40. DOI: 10.1023/A:1026740726955

[9] Dummett, M., "Victor's error", *Analysis*, 61 (2001): 1–2. DOI: 10.1093/analys/61.1.1

[10] Dummett, M., "Fitch's paradox of knowability", pages 51–52 in [27], Oxford University Press, Oxford, 2009. DOI: 10.1093/acprof:oso/9780199285495.003.0005

[11] Edington, D., "The paradox of knowability", *Mind*, 94 (1985): 557–568. DOI: 10.1093/mind/XCIV.376.557

[12] Gottwald, S., *A treatise on many-valued logic*, Baldock, Research Studies Press, 2001.

[13] Hintikka, J., *Knowledge and Belief*, Cornell University Press, Ithaca, N. Y., 1962.

[14] Kleene, S. C., "On a notation for ordinal numbers", *Journal of Symbolic Logic*, 3 (1938): 150–155. DOI: 10.2307/2267778

[15] Kleene, S. C., *Introduction to Metamathematics*, Van Nostrand, Amsterdam and Princeton, 1952.

[16] Łukasiewicz, J., "Philosophische Bemerkungen zu mehrwertigen Systemen des Aussagenkalküls", *Comptes rendus de la Société des Sciences et des Lettres de Varsovie*, 23 (1930): 1–21. English translation in [18].

[17] Łukasiewicz, J., and A. Tarski, "Untersuchungen über den Aussagenkalküls", *Comptes rendus de la Société des Sciences et des Lettres de Varsovie*, 23 (1930): 1–21. English translation in [18].

[18] Łukasiewicz, J., *Selected Works*, L. Borkowski (ed.), North-Holland, Amsterdam, 1970.

[19] Maffezioli, P., A. Naibo, A., and S. Negri, "The Church–Fitch knowability paradox in the light of structural proof theory", *Synthese*, 190, 14 (2013): 2677–2716. DOI: 10.1007/s11229-012-0061-7

[20] Nozick, R., *Philosophical Explanations* (Chapter 3), Harvard University Press, Cambridge, MA, 1981.

[21] Martin-Löf, P., "Truth and knowability: On the principles C and K of Michael Dummett", pages 105–114 in *Truth in mathematics*, G. Dales and G. Oliveri (eds.), Oxford University Press, Oxford, 1998.

[22] Odintsov, S. P., and H. Wansing, "The logic of generalized truth values and the logic of bilattices", *Studia Logica*, 103, 1 (2015): 91–112. DOI: 10.1007/s11225-014-9546-3

[23] Post, E., "Introduction to a general theory of elementary propositions", *American Journal of Mathematics*, 43 (1921): 163–185. DOI: 10.2307/2370324

[24] Priest, G., "Beyond the limits of knowledge", pages 93–104 in [27], 2009. DOI: 10.1093/acprof:oso/9780199285495.003.0008

[25] Proietti C., and G. Sandu, "Fitch's paradox and *ceteris paribus* modalities", *Synthese*, 173, 1, (2010): 75–87. DOI: 10.1007/s11229-009-9677-7

[26] Restall G., "Not every truth can be known (at least, not all at once)", pages 339–354, in [27], 2009. DOI: 10.1093/acprof:oso/9780199285495.003.0022

[27] Salerno J., *New Essays on the Knowability Paradox*, Oxford University Press, 2009.

[28] Shramko, Y., J. M. Dunn, and T. Takenaka, "The trilatice of constructive truth values", *Journal of Logic and Computation*, 11 (2001): 761–788. DOI: 10.1093/logcom/11.6.761

[29] Shramko, Y., and H. Wansing, "Some useful 16-valued logics: How a computer network should think", *Journal of Philosophical Logic*, 34, 2 (2005): 121–153. DOI: 10.1007/s10992-005-0556-5

[30] Shramko, Y., and H. Wansing, "Hyper-contradictions, generalized truth-values and logics of truth and falsehood", *Journal of Logic, Language and Information*, 15, 4 (2006): 403–424. DOI: 10.1007/s10849-006-9015-0

[31] Shramko, Y., and H. Wansing, *Truth and Falsehood. An Inquiry into Generalized Logical Values*, Springer, 2011.

[32] Tennant, N., *The Taming of the True*, Oxford University Press, Oxford, 1997.

[33] Tennant, N., "Revamping the restriction strategy", pages 223–238 in [27], 2009. DOI: 10.1093/acprof:oso/9780199285495.003.0015

[34] Wansing H., "Diamonds are a philosopher's best friends", *Journal of Philosophical Logic*, 31, 6 (2002): 591–612. DOI: 10.1023/A:1021256513220

[35] Williamson, T., "Intuitionism disproved?", *Analysis*, 42 (1982): 203–207. DOI: 10.1093/analys/42.4.203

[36] Williamson, T., "Verificationism and non-distributive knowledge", *Australasian Journal of Philosophy*, 71 (1993): 78–86. DOI: 10.1080/00048409312345072

[37] Wintein, S., and R. A. Muskens, "From bi-facial truth to bi-facial proofs", *Studia Logica*, 103, 3, (2015): 545–558. DOI: 10.1007/s11225-014-9578-8

[38] Zaitsev, D. V., “A few more useful 8-valued logics for reasoning with tetralattice EIGHT4”, *Studia Logica*, 92, 2 (2009): 265–280. DOI: 10.1007/s11225-009-9198-x

[39] Zaitsev, D. V., and Y. Shramko, “Bi-facial truth: A case for generalized truth values”, *Studia Logica*, 101, 6 (2013): 299–318. DOI: 10.1007/s11225-013-9534-z

[40] Zaitsev D., “Logics of generalized classical truth values”, pages 331–341 in *The Logica Yearbook 2014*, P. Arazim and M. Peliš (eds.), College Publications London, 2015.

Ekaterina Kubyshkina
Department of Philosophy
University of Paris 1 Panthéon – Sorbonne
Institute for History and Philosophy of Sciences and Technology
13, rue du Four,
75006 Paris, France
ekaterina.kubyshkina@univ-paris1.fr

Dmitry V. Zaitsev
Department of Philosophy
Lomonosov Moscow State University
“Shuvalovskiy” bldg, MSU, Leninskiye gory,
119991, Moscow, Russia
zaitsev@philos.msu.ru