STUDIES IN LOGIC, GRAMMAR
AND RHETORIC 62 (75) 2020
DOI: 10.2478/slgr-2020-0013

Arkadiusz Wójcik
University of Białystok
e-mail: a.wojcik@uwb.edu.pl
ORCID: 0000–0002–9895–8449

THE KNOWABILITY PARADOX AND UNSUCCESSFUL UPDATES

Abstract. In this paper we undertake an analysis of the knowability paradox in the light of modal epistemic logics and of the phenomena of unsuccessful updates. The knowability paradox stems from the Church-Fitch observation that the plausible knowability principle, according to which all truths are knowable, yields the unacceptable conclusion that all truths are known. We show that the phenomenon of an unsuccessful update is the reason for the paradox arising. Based on this diagnosis, we propose a restriction on the knowability principle which resolves the paradox.

Keywords: knowability paradox, knowability principle, dynamic epistemic logic, logic of public announcements, unsuccessful update, successful formulas.

1. Introduction

Providing a negative answer to a question about the existence of truths, the discovery of which lies entirely beyond human capabilities, is equivalent to stating the following theorem:

(The knowability principle) All truths are knowable.

The knowability principle is predominantly presented as a thesis adopted by the proponents of semantic anti-realism. However, a series of different philosophical concepts that are connected to accepting this thesis are mentioned in the literature. Joe Salerno (2009) attributes the acceptance of the knowability principle to, inter alia, logical positivists of the Vienna Circle, supporters of intuitionism, proponents of Kant's transcendental idealism, Charles Sanders Peirce's pragmatism and Hilary Putnam's internal realism. Otavio Bueno (2009) claims that the representatives of so-called "full-blooded platonism" (the view that all logically possible mathematical objects exist) as well as the proponents of mathematical fictionalism (see Field, 1980) are obliged to accept the knowability principle. Other authors – for reasons independent of the above-mentioned ideas – emphasise the intuitive acceptability of the knowability principle (see e.g. van Benthem, 2004; Beall, 2009; Priest, 2009; Quine, 1976; Restall, 2009). The "possibility" of knowing all truths mentioned in the principle is that of logical possibility. The existence of human cognitive limitations, as well as the limitations resulting from the laws of physics, is not an issue for those who want to embrace the knowability principle. It seems that due to such an understanding of the concept of possibility, the knowability principle should not be controversial.

However, the American logician and philosopher Frederic B. Fitch (1963) has shown that it is possible to prove – using a few non-controversial assumptions – that accepting this intuitively permissible knowability principle leads to the necessity of acknowledging the so-called omniscience principle:

(The omniscience principle) All truths are known.

The result achieved by Fitch is known as "the knowability paradox", Fitch's paradox of knowability or – on account of the role in Fitch's reasoning played by Alonzo Church's suggestion – "the Church-Fitch paradox". Since the reasoning presented by the American logician is not only noteworthy from a formal point of view, but also has significant philosophical implications, the knowability paradox soon became one prominent topic of discussion within the realms of philosophical logic and epistemology.

The sustained interest in the knowability paradox coincided with an unusually intensive development of epistemic logics, among which a central place is taken by modal systems built upon Saul Kripke's possible worlds semantics. Modal epistemic logics have become not only helpful tools for the formalization of certain intuitions connected with the concepts of knowledge, belief and information, but have also become a subject of interest to scientists in fields such as game theory, computer science, cognitive science, decision theory, artificial intelligence (AI) and cryptology. The increase of interest in epistemic logics among representatives of other scientific disciplines led to new goals for the application of the logic of knowledge and beliefs. Epistemic logics began to be perceived as formal systems whose aim is to capture the phenomenon of the epistemic change resulting from the flow of information between various agents. Thus, the need to develop formal systems that adequately capture the phenomenon of epistemic change led to a dynamic turnover in epistemic logic.

The phenomenon of an unsuccessful update is a particularly interesting phenomenon that can be successfully modelled in dynamic systems. This phenomenon states that (i) a certain true sentence becomes false as a consequence of its announcement and (ii) the announcing of this sentence alone leads to the negation of this sentence becoming known to all agents. The main purpose of this paper is to show that the phenomenon of an unsuccessful update is the reason for the knowability paradox arising. On the basis of this diagnosis of the paradox, we will propose a natural restriction of the knowability principle. This restriction will not generate paradoxical consequences and will avoid the charge of its being *ad hoc* – a charge that befalls other competing solutions.

2. The knowability paradox

The knowability principle is normally characterized in formal language in the following form:

(KP) $\varphi \to \Diamond K\varphi$,

where $\Diamond$ is a modal operator of possibility. Therefore $\Diamond\varphi$ should be read as "it is possible that φ". K is the knowledge operator, hence $K\varphi$ is interpreted as "it is known that φ" or "someone knows that φ".

The following assumptions are made in order to reach a paradoxical conclusion:

(A) $K(\varphi \wedge \psi) \to (K\varphi \wedge K\psi)$,

(B) $K\varphi \to \varphi$,

(C) if $\vdash \varphi$, then $\vdash \Box\varphi$,

(D) if $\vdash \Box\neg\varphi$, then $\vdash \neg\Diamond\varphi$.

The above-mentioned assumptions seem quite intuitive. According to (A), knowledge is distributed over conjunction, while on the basis of (B), knowledge implies truth. Rule (C) is a standard rule for all normal modal logics. This rule states that if a certain formula φ is provable, then formula $\Box\varphi$ is also provable, where $\Box\varphi$ is interpreted as "it is necessary that φ". Rule (D) is also not controversial, since $\Box$ and $\Diamond$ are mutually definable.

The reasoning which leads to the apparent paradox begins with substituting for (KP) the so-called Moore sentence $p \wedge \neg Kp$:

(1) $$(p \wedge \neg Kp) \to \Diamond K(p \wedge \neg Kp)$$

It is sufficient to consider the following reasoning in order to show that the Moore sentence cannot be the object of knowledge:

(2)	$K(p \wedge \neg Kp)$	premise for reductio
(3)	$K(p \wedge \neg Kp) \rightarrow (Kp \wedge K\neg Kp)$	substitution for (A)
(4)	$Kp \wedge K\neg Kp$	from (2) and (3) by classical logic
(5)	Kp	from (4) by classical logic
(6)	$K\neg Kp$	from (4) by classical logic
(7)	$K\neg Kp \rightarrow \neg Kp$	substitution for (B)
(8)	$\neg Kp$	from (6) and (7) by classical logic
(9)	$Kp \wedge \neg Kp$	from (5) and (8) by classical logic

We end in a contradiction. As a result of this reasoning, we can infer the negation of the premise from a *reductio ad absurdum* with (2):

$$\neg K(p \wedge \neg Kp) \tag{10}$$

In the next step we will use rule (C) applied to (10).

$$\Box\neg K(p \wedge \neg Kp) \tag{11}$$

Whereas we obtain from rule (D) applied to (11):

$$\neg\Diamond K(p \wedge \neg Kp) \tag{12}$$

Now, by applying modus tollens to (1) and (12), we get:

$$\neg(p \wedge \neg Kp) \tag{13}$$

Finally, in classical logic, from (13) we obtain:

$$p \rightarrow Kp \tag{14}$$

In the preceding line of reasoning, p could be substituted by any formula, which eventually leads to the conclusion:

$$\varphi \rightarrow K\varphi \tag{15}$$

Therefore, holding (KP) as well as the assumptions (A), (B), (C), and (D), results in the necessity of the omniscience principle, that is, of the theorem stating that all truths are already known.

At least several hundred papers have been published about the knowability paradox over the more than 50 years that have passed since the publication of Fitch's article. On the basis of this abundant literature it is possible to distinguish the following standard reactions to Fitch's paradox:

(i) rejecting one of the assumptions for the knowledge operator;

(ii) revising classical logic and adopting some non-classical logic;

(iii) restricting the knowability principle.

Employing strategy (i) seems the least credible – as already mentioned, it is difficult to undermine the justification of assuming (A) and (B). Besides, modified versions of Fitch's proof exist which do not use those assumptions or use their weaker versions (see e.g. Jago, 2010; Mackie, 1980; Williamson, 1993). Strategy (ii) entails adopting intuitionistic logic (see e.g. DeVidi & Solomon, 2001; Dummett, 2009; Williamson, 1988), paraconsistent logic (see Beall, 2000; Priest, 2009) or relevance logic (see Wansing, 2002). However, revising classical logic in answer to the knowability paradox is an *ad hoc* procedure. The proponents of semantic anti-realism can defend themselves against this objection because they adopt intuitionistic logic due to reasons independent from Fitch's paradox. In their case, the problematic reasoning stops at step (13) because the derivation of (14) from (13) calls for using the law of double negation. Thus the proponent of semantic anti-realism is obliged to accept the following consequence of Fitch's reasoning:

(15')
$$\varphi \to \neg\neg K\varphi$$

Yet, if negation is interpreted in an intuitionistic way, then not only does this formula not seem paradoxical, but it also truly reflects the essence of semantic anti-realism (see DeVidi & Solomon, 2001; Dummett, 2009). Nevertheless, other proponents of accepting the knowability principle would have to present additional arguments, independent from Fitch's paradox, for adopting non-classical logic in order to defend themselves from the allegation of an *ad hoc* solution.

The allegation of an *ad hoc* solution also applies to many examples of restricting the knowability principle to sentences with a certain property. The most highly debated proposition of this type was put forward by Neil Tennant (1997). Tennant has proposed to restrict the scope of the knowability principle to so-called Cartesian statements which are defined in the following way:

$$\varphi \text{ is Cartesian} \overset{\text{def}}{=} \neg(K\varphi \vdash \bot).$$

Thus the statement φ is Cartesian, if $K\varphi$ is consistent. Of course $p \wedge \neg Kp$ is not Cartesian, which was proved in steps (2)–(9) of the above-mentioned reasoning. However, Tennant did not put forward any arguments for the proposed restriction of the knowability principle except the fact that the proposed solution blocks the occurrence of paradox. Establishing a clear demarcation line between the *ad hoc* solutions and those that do not have that shortcoming is a difficult task. It seems, however, that to avoid the charge of having proposed an *ad hoc* solution, one should either put forward an argument independent from Fitch's paradox for one's proposed restriction of the knowability principle, or one's explanation of the paradox should properly justify one's proposed solution.

3. The dynamic epistemic logic

We will employ the formalism of dynamic epistemic logic in order to formulate a diagnosis of Fitch's paradox, on the basis of which a restriction of the knowability principle can be proposed.

Definition 1. Let Var be the set of propositional variables and let A be the set of agents. The language of dynamic epistemic logic $\mathcal{L}_{DEL}$ is defined inductively as follows:

$$\varphi \ ::= \ p \ | \ \neg\varphi \ | \ \varphi \rightarrow \varphi \ | \ K_a\varphi \ | \ [\varphi]\varphi$$

where $p \in Var$ and $a \in A$. The set of all $\mathcal{L}_{DEL}$ formulas is denoted by $\Gamma_{\mathcal{L}_{DEL}}$.

$\mathcal{L}_{DEL}$ is an extension of the language of propositional logic with epistemic operators K_a for every agent $a \in A$ as well as the dynamic operator $[\varphi]\psi$ for any formulas $\varphi, \psi \in \Gamma_{\mathcal{L}_{DEL}}$. The intended interpretation of $K_a\varphi$ is "agent a knows that φ", while the narrow interpretation of the formula $[\varphi]\psi$ is "after a public announcement φ, it holds that ψ". However, the default interpretation of the dynamic operator is wider, and $[\varphi]\psi$ should be read as "after an epistemic update with φ, it holds that ψ" or simply, "after obtaining information φ, it holds that ψ". Thus, one may get the updated epistemic model not only as a result of public announcements but also as a result of other acts, such as observation or positive verification. Depending on context, the formula $[\varphi]\psi$ can be interpreted, for example, as "after the observation that φ, it holds that ψ" or "after publicly verifying the truth of φ, it holds that ψ". Other classical logical constants can be defined in the standard way.

The semantics for $\mathcal{L}_{DEL}$ is constructed on the basis of the semantics that Kripke (1959, 1963) proposed for the logic of possibility and necessity.

Definition 2. An epistemic model is a structure $\mathcal{M} = (W, \{R_a : a \in A\}, v)$, where

$W \neq \emptyset$ is a set of epistemic states,

$A \neq \emptyset$ is a set of agents,

$R_a \subseteq W \times W$ is an epistemic accessibility relation for any $a \in A$,

$v : Var \mapsto \mathcal{P}(W)$ is a valuation function which to every $p \in Var$ assigns the set of epistemic states in which p is true.

When we consider standard epistemic modal logics (see e.g. Fagin et al., 1995), we deal with static semantics. Logics of this kind do not allow for modeling the phenomenon of a change in the agent's knowledge when new information is received. Dynamic logics enable us to model this phenomenon. The main idea involved in modeling updated knowledge is that whenever an agent receives new information, all epistemic states that are contradictory with this new information are removed from the epistemic model representing the agent's knowledge before the new information was received. Therefore the new information causes a transition from a certain initial epistemic model to its sub-model, that is, a model bounded by this new information.

Definition 3. Let $\mathcal{M} = (W, \{R_a : a \in A\}, v)$ be an epistemic model. The updated model by the new information φ is defined as a tuple $\mathcal{M}|\varphi = (W', \{R'_a : a \in A\}, v')$, where:

$W' = \{s \in W : \mathcal{M}, s \models \varphi\}$,

$R'_a = R_a \cap (W' \times W')$, for any $a \in A$,

$v'(p) = v(p) \cap W'$, for any $p \in Var$.

In other words, the model $\mathcal{M}|\varphi$ is the model $\mathcal{M}$ restricted to all those epistemic states where φ holds.

The introduced definitions allow for interpreting the language $\mathcal{L}_{DEL}$ in epistemic models, that is, formulating precise conditions for the satisfaction relation $\models$.

Definition 4. Let $\mathcal{M} = (W, \{R_a : a \in A\}, v)$ be an epistemic model. The satisfaction relation $\models$ is defined inductively in the following way:

$\mathcal{M}, s \models p$ iff $s \in v(p)$,

$\mathcal{M}, s \models \neg\varphi$ iff $\mathcal{M}, s \not\models \varphi$,

$\mathcal{M}, s \models \varphi \to \psi$ iff if $\mathcal{M}, s \models \varphi$, then $\mathcal{M}, s \models \psi$,

$\mathcal{M}, s \models K_a\varphi$ iff for all $t \in W$: if $(s, t) \in R_a$, then $\mathcal{M}, t \models \varphi$,

$\mathcal{M}, s \models [\varphi]\psi$ iff if $\mathcal{M}, s \models \varphi$, then $\mathcal{M}|\varphi, s \models \psi$,

where $p \in Var$, $\varphi, \psi \in \Gamma_{\mathcal{L}_{DEL}}$, $a \in A$, and $\mathcal{M}|\varphi$ is a model $\mathcal{M}$ restricted to all those epistemic states where φ holds.

The language $\mathcal{L}_{DEL}$ can be extended by the operator $\langle\ \rangle$, defined for any formulas $\varphi, \psi \in \Gamma_{\mathcal{L}_{DEL}}$ using the dynamic operator $[\ ]$ and negation:

$$\langle\varphi\rangle\,\psi \stackrel{\text{def}}{=} \neg[\varphi]\neg\psi.$$

Then, applying Definition 4, we get the following condition:

$\mathcal{M}, s \models \langle\varphi\rangle\,\psi$ iff $\mathcal{M}, s \models \varphi$ and $\mathcal{M}|\varphi, s \models \psi$.

Other important semantic terms are defined in the standard way. A formula $\varphi \in \Gamma_{\mathcal{L}_{DEL}}$ is true in an epistemic model $\mathcal{M} = (W, \{R_a : a \in A\}, v)$ whenever for any $s \in W$, $\mathcal{M}, s \models \varphi$. We denote this by $\mathcal{M} \models \varphi$. A formula $\varphi \in \Gamma_{\mathcal{L}_{DEL}}$ is valid whenever for any epistemic model $\mathcal{M}$, $\mathcal{M} \models \varphi$. We denote this by $\models \varphi$.

The discussed logic is obtained from the modal epistemic logic S5 by adding certain axiom schemes and inference rules.

Definition 5. A proof system for the logic DEL is given by the following axiom schemes and inference rules:

all instantiations of propositional tautologies

$K_a(\varphi \to \psi) \to (K_a\varphi \to K_a\psi)$

$K_a\varphi \to \varphi$

$K_a\varphi \to K_aK_a\varphi$

$\neg K_a\varphi \to K_a\neg K_a\varphi$

$[\varphi]p \leftrightarrow (\varphi \to p)$

$[\varphi]\neg\psi \leftrightarrow (\varphi \to \neg[\varphi]\psi)$

$[\varphi](\psi \wedge \chi) \leftrightarrow ([\varphi]\psi \wedge [\varphi]\chi)$

$[\varphi]K_a\psi \leftrightarrow (\varphi \rightarrow K_a[\varphi]\psi)$

$[\varphi][\psi]\chi \leftrightarrow [\varphi \wedge [\varphi]\psi]\chi$

The inference rules:

from $\varphi \rightarrow \psi$ and φ infer ψ (modus ponens)

if $\vdash \varphi$, then $\vdash K_a\varphi$ (Gödel's rule for K_a)

if $\vdash \varphi$, then $\vdash [\psi]\varphi$ (Gödel's rule for [ ]).

Jan Plaza (1988) as well as Jelle Gerbrandy and Willem Groeneveld (1997) independently proved the soundness and completeness theorem for the DEL logic.

Theorem 1. Logic **DEL** is sound and complete with respect to the class of all equivalence epistemic models, that is, any formula $\varphi \in \Gamma_{\mathcal{L}_{DEL}}$ is an **DEL**-theorem iff φ is true in all epistemic models, where the accessibility relations are the equivalence relations.

4. The Muddy Children Puzzle

The phenomenon of epistemic change is best illustrated by the so-called Muddy Children Puzzle, which comes from the book written by John E. Littlewood (1953).

Example. Three children come back home after playing in the garden. During the game, children could get mud on their foreheads. Each child sees the foreheads of the other two, however no child can see his or her own forehead. Their father waits for the children at home. He arranges the children in a row, and then says:

Father: At least one of you has a dirty forehead.

After this announcement he asks:

Father: Do any of you know whether your forehead is dirty? If so, then take a step forward.

None of the children step forward. Father repeats his question. Again nothing happens. Yet, when the father repeats the question for a third time, all children immediately take two steps forward.

Let us assume that we have the following trio of children: a, b, c. We will use propositional variables p_a, p_b, p_c to denote that relevant children have dirty foreheads. The graphical representation of the initial model of this situation is shown in Figure 1.

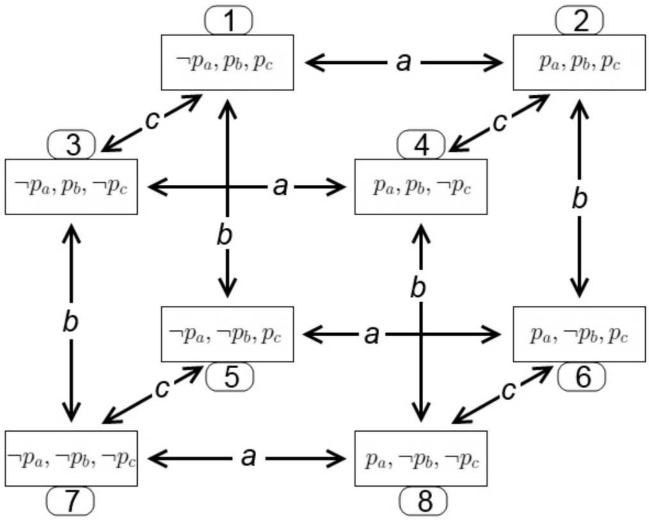

Figure 1. $\mathcal{M}$ (initial model)

There are eight possible worlds (epistemic states) here; for example, in world 1, we have $\neg p_a$, p_b, p_c, which means that in world 1 the forehead of child a is clean, while the foreheads of children b and c are dirty. The arrows represent the epistemic accessibility relation; for example, an arrow with the letter "a" leading from world 1 to world 2 and from world 2 to world 1 means that from the point of view of and according to the knowledge of child a, worlds 1 and 2 are indistinguishable. The relation of accessibility in this model is reflexive, but the arrows symbolising this fact are left out from the figures in order to achieve better graphical clarity.

The sequence of epistemic interactions begins from the father's first announcement: "At least one of you has a dirty forehead", which means that $p_a \vee p_b \vee p_c$ holds. Let us note that $\mathcal{M}, 7 \not\models p_a \vee p_b \vee p_c$. In the case of this puzzle, it is assumed that all information conveyed by agents is true information and that agents immediately notice all logical consequences of newly introduced information. Thus we can assume that all children eliminate world 7 as contradictory to the information announced by the father. The update of model $\mathcal{M}$ leads to its submodel $\mathcal{M}'$. The graphical representation of submodel $\mathcal{M}'$ is shown in Figure 2.

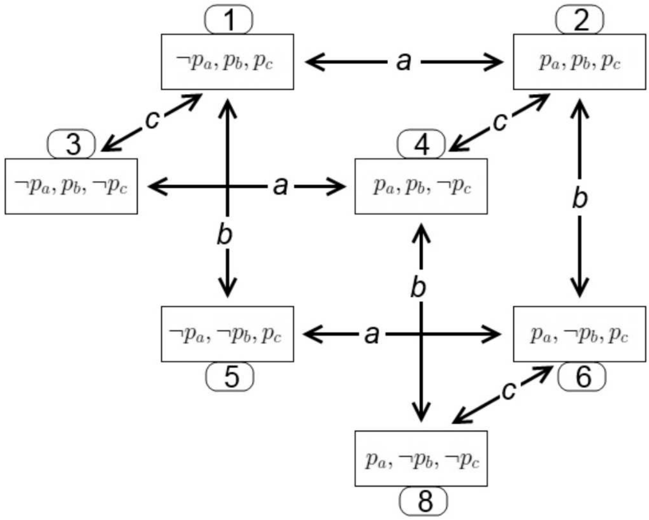

Figure 2. Model $\mathcal{M}'$ (model after Father's first announcement)

In the next step the father asks the children who know that they have a dirty forehead to take a step forward. Because none of the children take a step forward, agents receive new information: $\neg(K_a p_a \vee K_a \neg p_a) \wedge \neg(K_b p_b \vee K_b \neg p_b) \wedge \neg(K_c p_c \vee K_c \neg p_c)$. Let us note that $\mathcal{M}', 3 \models K_b p_b$, $\mathcal{M}', 5 \models K_c p_c$ and $\mathcal{M}', 8 \models K_a p_a$. Thus worlds 3, 5, and 8 are contradictory to the information received by children and are eliminated by the reasoning agents. The updated model's graphical representation is shown in Figure 3.

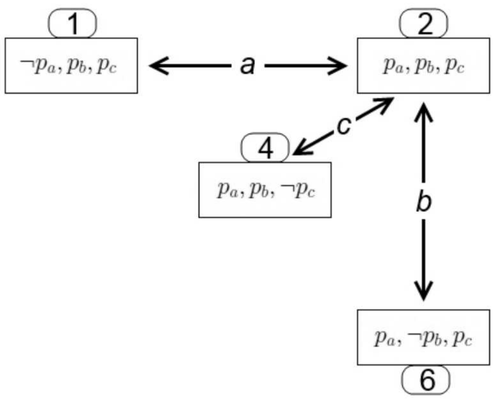

Figure 3. Model $\mathcal{M}''$ (model after the lack of reaction to Father's first question)

After the second question, the children also stay in place. Once more the children are provided with information that $\neg(K_a p_a \vee K_a \neg p_a) \wedge \neg(K_b p_b \vee$

$K_b\neg p_b) \wedge \neg(K_c p_c \vee K_c \neg p_c)$. Let us note that $\mathcal{M}'', 1 \models K_b p_b \wedge K_c p_c$, $\mathcal{M}'', 4 \models K_a p_a \wedge K_b p_b$ and $\mathcal{M}'', 6 \models K_a p_a \wedge K_c p_c$. Therefore worlds 1, 4 and 6 are contradictory to the introduced information and are eliminated. Thus the only world that is consistent to the introduced information is world 2, which is illustrated by Figure 4.

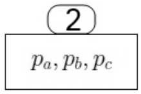

Figure 4. Final model (model after the lack of reaction to the father's second question)

Now the father can ask the question for a third time and all the children will step forward, since all already know that they have dirty foreheads.

5. The phenomenon of an unsuccessful update

The above example dealt with a specific situation – the fact that the children originally conveyed a lack of knowledge as to whether they had dirty foreheads led to them acquiring the knowledge that their foreheads are dirty. We call this type of phenomenon "an unsuccessful update". An unsuccessful update happens when a true sentence becomes false as a result of its announcement. The definition below introduces the distinction between successful and unsuccessful formulas and between successful and unsuccessful updates.

Definition 6. Let $\mathcal{M} = (W, \{R_a : a \in A\}, v)$ be an epistemic model and let $\varphi \in \Gamma_{\mathcal{L}_{DEL}}$.

φ is a successful formula iff $[\varphi]\varphi$ is valid in the DEL logic,

φ is an unsuccessful formula iff φ is not a successful formula,

φ is a successful update in state $s \in W$ iff $\mathcal{M}, s \models \langle\varphi\rangle\,\varphi$,

φ is an unsuccessful update in state $s \in W$ iff $\mathcal{M}, s \models \langle\varphi\rangle\,\neg\varphi$.

If φ is a successful (unsuccessful) update in state s, then we also claim that the update with φ is successful (unsuccessful) in state s or that φ leads to the phenomenon of successful (unsuccessful) update in state s.

It is necessary to explain the fact that the operator $[\ ]$ is used in the definition of successful and unsuccessful formulas, while the operator $\langle\ \rangle$ occurs in the definition of successful and unsuccessful updates. Let us note that if we accept the definition according to which φ is a successful formula if and only if $\langle\varphi\rangle\,\varphi$ is DEL-valid, it would mean that φ is a successful formula whenever for any $\mathcal{M}, s$ it is the case that $\mathcal{M}, s \models \varphi$ and $\mathcal{M}|\varphi, s \models \varphi$. This condition, however, can be satisfied only and exclusively for formulas that are tautologies. This way of defining successful formulas, then, will then have an undesirable consequence: all formulas that are not DEL-valid would be unsuccessful. Additionally, it is necessary to use the operator $\langle\ \rangle$ in the definition of successful and unsuccessful updates. Because if we assume that φ is an unsuccessful update in state s of model $\mathcal{M}$ whenever $\mathcal{M}, s \models [\varphi]\neg\varphi$ holds, then each false formula in a given state will be an unsuccessful update in it. However, this formulation does not adequately capture the phenomenon of an unsuccessful update. This phenomenon states that a true sentence becomes false as a consequence of its announcement.

On the one hand, from Definition 6 and how the operator $\langle\ \rangle$ was defined above, we get the result that updates from true and successful formulas are always successful. The reason for this is that a successful formula φ being true in state s of model $\mathcal{M}$, i.e. $\mathcal{M}, s \models \varphi \wedge [\varphi]\varphi$ holds, is equivalent to $\mathcal{M}, s \models \langle\varphi\rangle\,\varphi$, i.e. φ is a successful update in state s. On the other hand, sometimes updates with unsuccessful formulas are successful. The discussed example of Muddy Children describes exactly such a situation. Let us go back to this example and assume the following notation:

$$\psi := \neg(K_a p_a \vee K_a \neg p_a) \wedge \neg(K_b p_b \vee K_b \neg p_b) \wedge \neg(K_c p_c \vee K_c \neg p_c).$$

Let us note that e.g. $\mathcal{M}', 1 \not\models [\psi]\psi$, that is formula $[\psi]\psi$ is not valid in the DEL logic, therefore ψ is an unsuccessful formula. However, we have $\mathcal{M}', 2 \models \langle\psi\rangle\,\psi$, that is, in state 2 formula ψ is a successful update.

Given the semantic characteristic of successful formulas, it seems justified to ask how we should syntactically characterise this class of formulas. First, let us note certain negative result which shows that is not possible to define a class of successful formulas using a standard inductive definition.

Proposition 1.

(i) there is a formula $\varphi \in \Gamma_{\mathcal{L}_{DEL}}$ such that φ is successful, but formula $\neg\varphi$ is unsuccessful,

(ii) there are formulas $\varphi, \psi \in \Gamma_{\mathcal{L}_{DEL}}$ such that φ and ψ are successful, but formula $\varphi \wedge \psi$ is unsuccessful,

(iii) there are formulas $\varphi, \psi \in \Gamma_{\mathcal{L}_{DEL}}$ such that φ and ψ are successful, but formula $[\varphi]\psi$ is unsuccessful.

Proof. To begin with, we will indicate formulas with corresponding properties. Later in this paper, we will prove that formula $p \wedge \neg K_a p$ is unsuccessful. It is also possible to effortlessly prove that formula $\neg p \vee K_a p$ is successful (which is enough for proof (i)), as well as that formulas p and $\neg K_a p$ are successful (and hence we achieve proof (ii)). In order to justify the validity of (iii), it is enough to note that formulas p and $q \wedge \neg q$ are successful, while formula $[p](q \wedge \neg q)$ is unsuccessful. □

Although the presentation of general syntactic characteristics of successful formulas is still an open question, it should be noted that van Ditmarsch and Kooi (2006) have achieved certain partial results in this matter. For, the authors have defined a fragment of language $\mathcal{L}_{DEL}$ restricted to so-called positive formulas, that is, formulas in which for all $a \in A$ negation does not precede K_a. Following this, they have proved that all positive formulas are successful formulas.

6. Moore's sentences and unsuccessful updates

In the context of the knowability paradox, formula $p \wedge \neg K_a p$ is crucial. This formula – as featured in Fitch's argument – cannot be the object of knowledge of an agent to whom epistemic operator K_a refers. We will prove that this formula is an unsuccessful formula.

Proposition 2. Formula $p \wedge \neg K_a p$ is unsuccessful.

Proof. We will prove that formula $[p \wedge \neg K_a p]p \wedge \neg K_a p$ is not valid in the DEL logic. Let $\mathcal{M} = (W, R_a, v)$ be a model such that $W = \{s, t\}$, $R_a = \{(s, s), (t, t), (s, t), (t, s)\}$ and $v(p) = \{s\}$. Since $v(p) = \{s\}$, then $\mathcal{M}, s \models p$ and $\mathcal{M}, t \not\models p$. Since $\mathcal{M}, t \not\models p$ and $(s, t) \in R_a$, then $\mathcal{M}, s \not\models K_a p$. Therefore, $\mathcal{M}, s \models p \wedge \neg K_a p$. $\mathcal{M}|p \wedge \neg K_a p = (W', R'_a, v')$ is a submodel of model $\mathcal{M}$ such that $W' = \{s\}$, $R'_a = \{(s, s)\}$, $v'(p) = \{s\}$. Let us note that $\mathcal{M}|p \wedge \neg K_a p, s \models K_a p$, which implies that $\mathcal{M}|p \wedge \neg K_a p, s \not\models p \wedge \neg K_a p$. Finally, we have $\mathcal{M}, s \models p \wedge \neg K_a p$ and $\mathcal{M}|p \wedge \neg K_a p, s \not\models p \wedge \neg K_a p$, therefore $\mathcal{M}, s \not\models [p \wedge \neg K_a p]p \wedge \neg K_a p$. This proves that formula $[p \wedge \neg K_a p]p \wedge \neg K_a p$ is not DEL-valid. □

It should also be noted that in model $\mathcal{M}$ described above, $\mathcal{M}, s \models \langle p \wedge \neg K_a p \rangle \neg(p \wedge \neg K_a p)$ holds, therefore in state s formula $p \wedge \neg K_a p$ is an unsuccessful update.

As the example below shows, an occurrence of an unsuccessful update due to the announcement $p \wedge \neg K_a p$ is not extremely rare, but can happen in communicative situations we encounter every day.

Example (**An unsuccessful update**). Agent A eagerly waits for a university acceptance letter. Agent B, who knows that A waits for this information, picks up the letter from the post office, reads its contents and finds out that A has been accepted. Then B calls A and says the following sentence:

(S) You do not know, but you have been accepted to university.

B conveyed two pieces of information in his announcement: (i) A has been accepted to university; (ii) A does not know that A has been accepted to university. Let us assume that agent A considers B a reliable source of information. Conveying piece of information (i) renders piece of information (ii) false. Therefore, true sentence (S), due to its announcement, became false.

We propose an extension of the analysis presented so far of formulas from the language $\mathcal{L}_{DEL}$ and we distinguish another class of formulas: strictly unsuccessful formulas. Strictly unsuccessful formulas are always (i.e. in each state of any epistemic model) false after their announcement.

Definition 7. Let $\varphi \in \Gamma_{\mathcal{L}_{DEL}}$. Formula φ is a strictly unsuccessful if and only if $[\varphi]\neg\varphi$ is valid in the DEL logic.

Introducing this distinction is justified because not all unsuccessful formulas are also strictly unsuccessful formulas.

Proposition 3. There is a formula $\varphi \in \Gamma_{\mathcal{L}_{DEL}}$ such that φ is unsuccessful, but it is not strictly unsuccessful.

Proof. Let us consider model $\mathcal{M}'$ the graphical representation of which is shown in Figure 2. Let us assume notation $\psi := \neg(K_a p_a \vee K_a \neg p_a) \wedge \neg(K_b p_b \vee K_b \neg p_b) \wedge \neg(K_c p_c \vee K_c \neg p_c)$. Since e.g. $\mathcal{M}', 1 \not\models [\psi]\psi$ holds, formula ψ is not successful. However, ψ is not a strictly unsuccessful formula, because it holds that $\mathcal{M}', 2 \models [\psi]\psi$. $\square$

Let us note that – in contrast to unsuccessful formulas – updates performed by true and strictly unsuccessful formulas are always unsuccessful: if a strictly unsuccessful formula φ is true in state s of model $\mathcal{M}$, i.e. $\mathcal{M}, s \models \varphi \wedge [\varphi]\neg\varphi$ holds, then according to Definition 4, $\mathcal{M}, s \models \langle\varphi\rangle \neg\varphi$ holds, therefore φ is an unsuccessful update.

How to formulate the syntactic characteristics of strictly unsuccessful formulas is an interesting open question. From the point of view of research into Fitch's paradox, however, the most important fact is that an example of strictly unsuccessful formula is formula $p \wedge \neg K_a p$.

Proposition 4. Formula $p \wedge \neg K_a p$ is strictly unsuccessful.

Proof. We will show that formula $[p \wedge \neg K_a p]\neg(p \wedge \neg K_a p)$ is valid in the logic DEL. Let $\mathcal{M} = (W, \{R_a : a \in A\}, v)$ be a model of the logic DEL and let $s \in W$. Let us assume that $\mathcal{M}, s \models p \wedge \neg K_a p$. Suppose for contradiction that $\mathcal{M}|p \wedge \neg K_a p, s \models p \wedge \neg K_a p$. Hence there is $t \in W'$ such that $(s, t) \in R'_a$ and $\mathcal{M}|p \wedge \neg K_a p, t \not\models p$, where $W' = \{x \in W : \mathcal{M}, x \models p \wedge \neg K_a p\}$, $R'_a = R_a \cap (W' \times W')$. However, since $t \in W'$, then $\mathcal{M}, t \models p$ and also $\mathcal{M}|p \wedge \neg K_a p, t \models p$. Contradiction. Therefore $\mathcal{M}|p \wedge \neg K_a p, s \not\models p \wedge \neg K_a p$. This proves that formula $[p \wedge \neg K_a p]\neg(p \wedge \neg K_a p)$ is DEL-valid. □

On the basis of Proposition 4, it is known that if formula $p \wedge \neg K_a p$ is true in a given state of specific epistemic model, then $p \wedge \neg K_a p$ is an unsuccessful update in this state, i.e., the announcement $p \wedge \neg K_a p$ itself changes the logical value of this formula. However, as we remarked above, the interpretation of the formula $[\varphi]\psi$ can be wider than the standard "after a public announcement φ, it holds that ψ". Moreover, the update of an epistemic model can also be performed due to acts other than public announcements, e.g., observations or positive verifications. The result obtained in Proposition 4 thus indicates that the update of an epistemic model with $p \wedge \neg K_a p$, that happens after a certain cognitive action, makes formula $p \wedge \neg K_a p$ false. This explains why $p \wedge \neg K_a p$ cannot be known by an agent a: for a to obtain knowledge that $p \wedge \neg K_a p$, an update of epistemic model with $p \wedge \neg K_a p$ must have happened, but at that time due to this update $p \wedge \neg K_a p$ would already have been false, and therefore – on the basis of axiom schema $K_a\varphi \rightarrow \varphi$ – could not become an object of knowledge for anyone. Moreover, after the update of an epistemic model by $p \wedge \neg K_a p$, the negation of this formula becomes known, because the following proposition holds:

Proposition 5. Let $\varphi \in \Gamma_{\mathcal{L}_{DEL}}$ and $a \in A$. Then $[\varphi]\neg\varphi$ is valid if and only if $[\varphi]K_a\neg\varphi$ is valid.

Proof. We will show that for the logic DEL it holds that if $\models [\varphi]\neg\varphi$, then $\models [\varphi]K_a\neg\varphi$. Let us assume that $\models [\varphi]\neg\varphi$. Let $\mathcal{M} = (W, \{R_a : a \in A\}, v)$

be a model of the logic DEL, let $s \in W$ and $a \in A$. Suppose that $\mathcal{M}, s \models \varphi$. We shall show that $\mathcal{M}|\varphi, s \models K_a\neg\varphi$. Thus let $t \in W'$ be such that $(s,t) \in R_a^{'}$, where $W' = \{x \in W : \mathcal{M}, x \models \varphi\}$ and $R_a' = R_a \cap (W' \times W')$. Since $t \in W'$, then $\mathcal{M}, t \models \varphi$. However, according to the assumption that $\models [\varphi]\neg\varphi$, we have $\mathcal{M}|\varphi, t \models \neg\varphi$. Therefore $\mathcal{M}|\varphi, s \models K_a\neg\varphi$. This prooves that $\models [\varphi]K_a\neg\varphi$. In order to prove the converse implication, it suffices to recall that in any DEL-model relation R_a is reflexive for any $a \in A$. □

Employing the formalism of dynamic logics allows us to show that the phenomenon of an unsuccessful update is the reason for the knowability paradox arising. Furthermore, unlike with competing solutions, it provides the materials for a restriction of the knowability principle that is not *ad hoc*. The condition we must place on the knowability principle is that it be restricted to successful formulas. As mentioned above, updates by true and successful formulas are always successful. Moreover, it is possible to prove that any successful formula becomes an object of knowledge after verifying that this formula is true.

Proposition 6. Let $\varphi \in \Gamma_{\mathcal{L}_{DEL}}$ and $a \in A$. Then $[\varphi]\varphi$ is valid if and only if $[\varphi]K_a\varphi$ is valid.

Proof. The proof is analogous to that of Proposition 5. □

7. Summary

Judging from the number of publications dedicated to the knowability paradox, it is fair to say that the reasoning presented by Fitch caused an upheaval in philosophical logic. Many logicians felt obliged to provide some kind of answer to the perplexing reasoning; some of these answers had a revolutionary character, like that of postulating a revision of classical logic. However, analysis of the paradox using the apparatus of dynamic epistemic logic shows that such radical reactions are not required. The falsity of the knowability principle is caused by the existence of the phenomenon of an unsuccessful update which states that a true sentence becomes false due to its verification. If we discard the delusion that all formulas and updates are successful, then the apparent paradoxicality will disappear. It may be more accurate to say that Fitch's *paradox* is rather Fitch's *observation*: not all unknown truths can be known. This observation, however, leads to an obvious question – how to restrict the knowability principle in order to avoid the allegation of proposing an *ad hoc* restriction. In this paper, we argued that the relevant restriction of the knowability principle should be to successful formulas, that is, those that do not change their logical value during the process of verification. Of course, the possibility of defining another class of formulas leading to successful updates – the proper subset of which will be successful formulas – is not ruled out. In that case, the knowability principle should be restricted to this more comprehensive class of formulas. However, as a result of the conceptualisation of the knowability problem that we provide in this paper, this matter acquires a more technical character and we gain new formal tools for examining it.

REFERENCES

Beall, J. C. (2009). Knowability and Possible Epistemic Oddities. In J. Salerno (Eds.), *New Essays on the Knowability Paradox* (pp. 105–125). New York: Oxford University Press.

van Benthem, J. (2004). What One May Come to Know. *Analysis*, *64*(2), 95–105.

Bueno, O. (2009). Fitch's Paradox and the Philosophy of Mathematics. In J. Salerno (Eds.), *New Essays on the Knowability Paradox* (pp. 252–280). New York: Oxford University Press.

DeVidi, D., & Solomon, G. (2001). Knowability and Intuitionistic Logic. *Philosophia*, *28*, 319–334.

van Ditmarsch, H., & Kooi, B. (2006). The Secret of My Success. *Synthese*, *151*(2), 201–232.

Dummett, M. (2009). Fitch's Paradox of Knowability. In J. Salerno (Eds.), *New Essays on the Knowability Paradox* (pp. 51–52). New York: Oxford University Press.

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

Field, H. (1980). *Science Without Numbers*. Princeton: Princeton University Press.

Fitch, F. B. (1963). A Logical Analysis of Some Value Concepts. *Journal of Symbolic Logic*, *28*(2), 135–142.

Gerbrandy, J., & Groeneveld, W. (1997). Reasoning About Information Change. *Journal of Logic, Language, and Information*, *6*(2), 147–169.

Jago, M. (2010). Closure on Knowability. *Analysis*, *70*(4), 648–659.

Kripke, S. (1959). Semantical Analysis of Modal Logic (abstract). *Journal of Symbolic Logic*, *24*(4), 323–324.

Kripke, S. (1963). Semantical Analysis of Modal Logic I. Normal Modal Propositional Calculi. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, *9*, 67–96.

Littlewood, J. E. (1953). *A Mathematician's Miscellany*. London: Meuthen.

Mackie, J. L. (1980). Truth and Knowability. *Analysis*, *40*(2), 90–92.

Plaza, J. (1989). Logics of Public Communications. In M. Emrich, M. Pfeifer, M. Hadzikadic, & Z. Ras (Eds.), *Proceedings of the 4th International Symposium on Methodologies for Intelligent Systems: Poster Session Program* (pp. 201–216). Oak Ridge: Oak Ridge National Laboratory.

Priest, G. (2009). Beyond the Limits of Knowledge. In J. Salerno (Eds.), *New Essays on the Knowability Paradox* (pp. 93–104). New York: Oxford University Press.

Quine W. V. (1976). The Limits of Knowledge. In W. V. Quine (Eds.), *The Ways of Paradox and Other Essays* (pp. 59–67). New York: Random House.

Restall, G. (2009). Not Every Truth Can Be Known (At Least, Not all at Once). In J. Salerno (Eds.), *New Essays on the Knowability Paradox* (pp. 339–354). New York: Oxford University Press.

Salerno, J. (2009). Introduction. In J. Salerno (Eds.), *New Essays on the Knowability Paradox* (pp. 1–10). New York: Oxford University Press.

Tennant, N. (1997). *The Taming of the True*. Oxford: Oxford University Press.

Wansing, H. (2002). Diamonds are a Philosopher's Best Friends. *Journal of Philosophical Logic*, *31*(6), 591–612.

Williamson, T. (1988). Knowability and Constructivism. *Philosophical Quarterly*, *38*(153), 422–432.

Williamson, T. (1993). Verificationism and Non-Distributive Knowledge. *Australasian Journal of Philosophy*, *71*(1), 78–86.