Provability logic meets the knower paradox

Mirjam de Vos, Barteld Kooi [1]

Department of Theoretical Philosophy, Faculty of Philosophy, University of Groningen
Oude Boteringestraat 52, 9712 GL Groningen, The Netherlands

Rineke Verbrugge [2]

Institute of Artificial Intelligence, Faculty of Science and Engineering, University of Groningen
PO Box 407, 9700 AK Groningen, The Netherlands

1 Introduction

The knower paradox revolves around the following statement:

We know that statement (P) is false. (P)

The paradoxical reasoning goes as follows: Suppose (P) is true. Since it states that 'we know statement (P) is false', it is true that we know statement (P) is false. Since knowledge implies truth (a usual epistemological assumption), it follows that statement (P) is false, thus leading to a contradiction. Hence statement (P) is false and since *we* were the ones who just proved it, we know it to be false. However, this is exactly what statement (P) states, therefore it is true. So we have arrived at a paradox.

The knower paradox was first formulated by Kaplan and Montague [6]. They used *elementary syntax*, by which they understood "a first-order theory containing (...) all standard names (of expressions), means for expressing syntactical relations between, and operations on, expressions, and appropriate axioms involving these notions" [6, Footnote 10, p. 89]. Note that by elementary syntax they meant both a formal language and some sort of proof system. Kaplan and Montague used '$\varphi \vdash \psi$' to express that ψ is derivable from φ within the theory and '$\vdash \varphi$' means that φ is provable within this theory. In addition, they used names for expressions, where $\overline{\varphi}$ denotes the name of expression φ. The following two formulae are added to the elementary syntax:

$$\begin{array}{ll} K(\overline{\varphi}) & A \text{ knows the expression } \varphi \\ I(\overline{\varphi}, \overline{\psi}) & \varphi \vdash \psi \end{array}$$

According to Kaplan and Montague [6, p. 87], we can now formalize (P), as follows:

$$\vdash D \leftrightarrow K(\overline{\neg D}).$$

From this expression, some version of the knower paradox is derived, if the following three assumptions are made:

[1] mirjam.a.de.vos@gmail.com, B.P.Kooi@rug.nl

[2] L.C.Verbrugge@rug.nl

$$E1 := K(\overline{\neg D}) \to \neg D \tag{E1}$$
$$E2 := K(\overline{E_1}) \tag{E2}$$
$$E3 := [I(\overline{E_1}, \overline{\neg D}) \wedge K(\overline{E_1})] \to K(\overline{\neg D}) \tag{E3}$$

We derive the knower paradox as follows:

(1)	$\vdash D \leftrightarrow K(\overline{\neg D})$	by definition of D
(2)	$\vdash D \to K(\overline{\neg D})$	by (1), PC
(3)	$E1 \vdash K(\overline{\neg D}) \to \neg D$	by definition of $E1$
(4)	$E1 \vdash D \to \neg D$	by (2), (3), HS
(5)	$E1 \vdash \neg D$	by (4), PC
(6)	$E1 \vdash I(\overline{E_1}, \overline{\neg D})$	(5), by definition of I
(7)	$E1, E2 \vdash K(\overline{E1})$	by definition of $E2$
(8)	$E1, E2 \vdash I(\overline{E_1}, \overline{\neg D}) \wedge K(\overline{E_1})$	by (6), (7), PC
(9)	$E1, E2, E3 \vdash [I(\overline{E_1}, \overline{\neg D}) \wedge K(\overline{E_1})] \to K(\overline{\neg D})$	by definition of $E3$
(10)	$E1, E2, E3 \vdash K(\overline{\neg D})$	by (8), (9), MP
(11)	$E1, E2, E3 \vdash K(\overline{\neg D}) \to D$	by (1), PC
(12)	$E1, E2, E3 \vdash D$	by (10), (11), MP
(13)	$E1, E2, E3 \vdash \neg D \to D$	by (12), PC
(14)	$E1, E2, E3 \vdash D \leftrightarrow \neg D$	by (4), (13), PC

Over the years, many solutions have been proposed. We take Paul Égré's paper [4] as our point of departure. He argues that the knower paradox is solvable when modal provability logic is applied and uses three different interpretations of provability logic to solve the paradox, of which we focus on the one inspired by Solovay [9]. Provability logic provides a natural system in which both modalities and self-reference are treated rigorously, hence it may shed some light on where the paradoxical reasoning underlying the paradox goes awry. Our main contribution is an assessment of how the interpretation by Solovay fares in the light of Susan Haack's criteria for solutions to paradoxes [5], which include both technical and philosophical desiderata. In this way we hope to advance the debate regarding the knower paradox.

2 Provability Logic and Formal Systems of Arithmetic

We give a short reminder of Peano arithmetic and the provability logics **GL** and **GLS**.

2.1 Robinson Arithmetic and Peano Arithmetic

Here follow the axioms of Robinson arithmetic Q [8]: $\forall x(0 \neq Sx)$; $\forall x\forall y(Sx = Sy \to x = y)$; $\forall x(x \neq 0 \to \exists y(x = Sy))$; $\forall x(x + 0 = x)$; $\forall x\forall y(x + Sy = S(x + y))$; $\forall x(x \cdot 0 = 0)$; $\forall x\forall y(x \cdot Sy = (x \cdot y) + x)$. A statement φ is a theorem of Q if it is (an instance of) an axiom or if it can be derived from the axioms by the available rules of inference, modus ponens and generalization. Peano arithmetic (PA) [7] extends Q by the *Induction Schema*:$\{\varphi(0) \wedge \forall x(\varphi(x) \to \varphi(Sx))\} \to \forall x\varphi(x)$.

Theorem 2.1 (Diagonal Lemma, [2, p. 54] [3] **)** *Suppose that* $P(y)$ *is a formula of the language of* PA *in which no variable other than* y *is free. Then there exists a sentence* S *of the language of* PA *such that* $\mathrm{PA} \vdash S \leftrightarrow P(\overline{S})$.

2.2 Provability Logic

The provability logic **GL** contains the following axioms:

$$\text{All (instances of) propositional tautologies} \tag{A1}$$
$$\Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi) \tag{A2}$$
$$\Box(\Box\varphi \to \varphi) \to \Box\varphi \tag{GL}$$

The rules of inference of **GL** are modus ponens and necessitation (if $\varphi \in \mathbf{GL}$, then $\Box\varphi \in \mathbf{GL}$). Note that $\Box\varphi \to \Box\Box\varphi \in \mathbf{GL}$ [10].

2.3 The Relation between Provability Logic and Peano Arithmetic

A *realization* is a function $*$ assigning to each propositional atom of modal logic a sentence of the language of arithmetic, inductively defined by: $\bot^* = \bot; (\varphi \to \psi)^* = (\varphi^* \to \psi^*); (\Box\varphi)^* = \mathit{Prov}(\overline{\varphi^*})$. Solovay [9] proved that **GL** is arithmetically complete with respect to PA. The arithmetical soundness of **GL** was already clear. Formally:

$$\mathbf{GL} \vdash \varphi \text{ if and only if } \mathrm{PA} \vdash \varphi^* \text{ for all realizations } * .$$

The system **GLS**, defined by Solovay [9, Section 5.1] [4] , contains all theorems of **GL** as axioms as well as all instances of the reflection principle $\Box\varphi \to \varphi$, and modus ponens is its single rule of inference. As for **GL**, the system **GLS** enjoys arithmetical soundness and completeness, but with respect to the standard model:

$$\mathbf{GLS} \vdash \varphi \text{ if and only if } \langle \omega; +, \cdot \rangle \models \varphi^* \text{ for all realizations } * .$$

Égré [4, p. 43] defines PA^+ as the closure under *modus ponens* of PA, supplemented with all instances of the reflection principle $\mathit{Prov}(\overline{A}) \to A$. PA^+ is stronger than PA because it can now prove the consistency of PA as an instance of reflection: $\mathit{Prov}(\overline{\bot}) \to \bot$. It is not hard to prove that we also have: [5]

$$\mathbf{GLS} \vdash \varphi \text{ if and only if } \mathrm{PA}^+ \vdash \varphi^* \text{ for all realizations } * .$$

3 Solving the knower paradox using provability logic

We consider Solovay's **GLS**, which solves the knower paradox according to Égré [4]. We then discuss whether the solution satisfies Haack's criteria [5].

Why is the knower paradox prevented in **GLS**? Remember that $K(\overline{E1})$, where $E1$ was defined as $K(\overline{\neg D}) \to \neg D$, was needed in the derivation of the knower paradox by Kaplan and Montague [6] (see page 2, Step (7)). In **GLS**, we have $\Box\neg D \to \neg D$ as an instance of the reflection principle. Because necessitation is not an inference rule of **GLS**, $\Box(\Box\neg D \to \neg D)$ cannot be derived from the reflection principle, therefore, Kaplan and Montague's derivation cannot be repeated in **GLS**. Let us now assess to which extent Solovay's **GLS** satisfies Haack's criteria for solutions to paradoxes.

---

[4] We follow current conventions as in e.g. [2] in that Solovay's G is our **GL** and his G' is our **GLS**.

[5] The proof is in our journal manuscript under revision, "Solutions to the knower paradox in the light of Haack's criteria".

3.1 The Formal Part of Solovay's Theory as a Solution

Haack's first requirement on solutions to paradoxes says that a solution should contain a consistent formal system indicating an unacceptable premise, principle of inference, or set of theorems [5]. Solovay's formal system **GLS** indicates the rejection of $K(\overline{K(\overline{\varphi}) \to \varphi})$, which is achieved by disallowing the necessitation rule to be applied to the reflection principle $K(\overline{\varphi}) \to \varphi$. Is **GLS** consistent? Solovay [9] proved that **GLS** is arithmetically sound with respect to the standard model. Since truth in a model implies consistency, **GLS** is consistent. So Haack's first requirement is satisfied.

3.2 The Philosophical Part of Solovay's Theory as a Solution

Haack's second requirement on solutions to paradoxes says that they should provide a philosophical explanation of why the suspect premise or principle of inference seems acceptable but is unacceptable. So in the case of **GLS**, there needs to be an argument for rejecting $K(\overline{K(\overline{\varphi}) \to \varphi})$ and/or for disallowing the necessitation rule to apply to the reflection principle $K(\overline{\varphi}) \to \varphi$. This argumentation should be independent of the existence of the knower paradox. Solovay [9] did not consider **GLS** within the context of the knower paradox. His article is about provability, not about knowledge, so it does not contain arguments for rejecting $K(\overline{K(\overline{\varphi}) \to \varphi})$ itself.

For provability, there are independent reasons to reject $Prov(\overline{Prov(\overline{\varphi}) \to \varphi})$. The formalized version of Löb's theorem states that $\mathrm{PA} \vdash Prov(\overline{Prov(\overline{\varphi}) \to \varphi}) \to Prov(\overline{\varphi})$. So if $Prov(\overline{Prov(\overline{\varphi}) \to \varphi})$ were accepted as an axiom scheme, then $Prov(\overline{\varphi})$ would hold for every statement φ, even for false statements.

Égré [4, p. 42] argues that **GL** can be seen as a "system formalizing the knowledge of an ideal mathematician recursively generating all the theorems of PA and reflecting on the scope of his knowledge". However, the only reason mentioned in [4] for accepting **GLS** is not independent of the existence of the paradox: the necessitation rule is not allowed to be applied to the reflection principle ($K(\overline{\varphi}) \to \varphi$) just to prevent the paradox. Therefore, for the solution to satisfy Haack's second requirement, we will need reasons to let a knowledge predicate satisfy the axioms of **GLS** independently of the knower paradox.

Let us therefore attempt to give an independent reason. Since **GLS** is arithmetically sound with respect to the standard model $\langle \omega; +, \cdot \rangle$, for every formula φ in the language of **GL**, $\mathbf{GLS} \vdash \Box\varphi$ implies $\omega \models Prov_{\mathrm{PA}}(\overline{\varphi^*})$, therefore there exists a proof of φ^* in PA, therefore $\mathrm{PA} \vdash \varphi^*$ for every realization $*$. So for $\Box$ interpreted as knowledge, **GLS** is *epistemically conservative* over PA, meaning that **GLS** will not prove any 'new' formulas of the form 'It is known that φ', i.e. $\Box\varphi$, for which Peano Arithmetic does not prove φ^* yet (cf. [3]). This is an argument to accept the theory **GLS** as a solution to the knower paradox.

3.3 The Scope of Solovay's Theory as a Solution

Haack's third requirement states that a solution to a paradox should not be too broad or too narrow. Solovay's system **GLS** is consistent, so it does not prove just everything. On the other hand, it does prove other desired theorems than just those needed to formalize the knower paradox. For example, Gödel sentence G in the language of PA

satisfies PA $\vdash G \leftrightarrow \neg Prov(\overline{G})$. Is there a sentence G satisfying **GLS** $\vdash G \leftrightarrow \neg\Box G$? Yes there is, namely $\neg\Box\bot$. The formula $\neg\Box\bot \leftrightarrow \neg\Box\neg\Box\bot$ is provable in **GL**, and thus in **GLS** (see [10, Section 2.2]). Therefore, **GLS** is neither too broad nor too narrow.

3.4 Discussion: Interpreting Knowledge as Provability

Why does it seem at first sight intuitively plausible to interpret knowledge (facts about Peano arithmetic known by some mathematicians) as provability in PA? According to many non-Platonists, proofs are constructed by mathematicians: there exists a proof of a certain statement only if there has been a mathematician who proved it. Thus, a statement can only be provable if it is already known: A statement that will be proved only next year is not provable yet.

However, Platonists can argue against interpreting knowledge as provability that knowledge depends on mathematicians and on time, while provability does not. There are statements known by mathematicians but not provable in PA, such as the Gödel sentence for PA. Conversely, a theorem provable in PA but for a long time not known is the formalized version of Löb's theorem.

Justification logic provides an interesting formal take on the correspondence between knowledge and provability. It includes *explicit knowledge* $t : \varphi$, meaning that 'φ is justified by t'. A logic combining implicit and explicit knowledge is **S4LP** [1], which contains the axiom scheme $(t : \varphi) \rightarrow \Box\varphi$. Since a proof is a form of justification, this justification logic is a way to connect knowledge and provability.

3.5 Conclusion

Summarizing the discussion about the quality of Solovay's system **GLS** as a solution to the knower paradox, Haack's first requirement is clearly satisfied and the solution (after our addition to [4]) satisfies the second criterion. The third criterion is provisionally met, because the solution is not too narrow and provisionally not too broad.

References

[1] Artemov, S. and E. Nogina, *Introducing justification into epistemic logic*, Journal of Logic and Computation **15** (2005), pp. 1059–1073.

[2] Boolos, G., "The Logic of Provability," Cambridge University Press, Cambridge, 1995.

[3] Dean, W. and H. Kurokawa, *The paradox of the Knower revisited*, Annals of Pure and Applied Logic **165** (2014), pp. 199–224.

[4] Égré, P., *The knower paradox in the light of provability interpretations of modal logic*, Journal of Logic, Language and Information **14** (2005), pp. 13–48.

[5] Haack, S., "Philosophy of Logics," Cambridge University Press, Cambridge, 1978.

[6] Kaplan, D. and R. Montague, *A paradox regained*, Notre Dame Journal of Formal Logic **1** (1960), pp. 79–90.

[7] Peano, G., "Arithmetices Principia, Nova Methodo Exposita," Fratres Bocca, Turin, 1889.

[8] Robinson, R., *An essentially undecidable axiom system*, in: *Proceedings of the International Congress of Mathematicians, Cambridge, Volume 1*, 1950, pp. 729–730.

[9] Solovay, R. M., *Provability interpretations of modal logic*, Israel Journal of Mathematics **25** (1976), pp. 287–304.

[10] Verbrugge, R., *Provability logic*, in: E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*, 2017, fall 2017 edition .