Three solutions to the knower paradox

Francesca Poggiolesi

Three Solutions to the Knower Paradox

Francesca Poggiolesi

In this paper I shall present three solutions to the Knower Paradox which, despite important points in common, differ in several respects. The first solution, proposed by C. A. Anderson [1] is a hierarchical solution, developed in the framework of first-order arithmetic. However I will try to show that this solution is based on an incorrect argument. The second solution, inspired by a book of R.M. Smullyan [14], is developed in the framework of modal logic and it is based on the idea of interpreting one of the basic systems of the modal logic of provability in an epistemic way. I shall give arguments in support of this solution. The third solution, proposed by P. Egrèé [8] attempts to connect the first and the second solutions. I will show that this attempt fails for philosophical and formal reasons.

Keywords: *hierarchy solutions, Knower Paradox, provability logic, self-reference.*

1. *Two versions of the Knower Paradox*

The Knower Paradox was introduced in 1960 by Kaplan and Montague [9] and basically is the epistemological counterpart of the Liar Paradox. I will start by presenting it in its original version, the one which treats knowledge as a predicate attached to names of sentences and uses the framework of first order arithmetic.

Let then T be a theory which extends Robinson Arithmetic (see, for example, [4], pp. 110) and satisfies the following three conditions.

First Condition. T can prove the fixed point lemma.

The fixed point lemma. For every formula P(x) of $\mathcal{L}$(T), with one free variable x, there exists a sentence α of $\mathcal{L}$(T) such that $T \vdash \alpha \leftrightarrow P(\overline{\alpha})$ (where the overlining of a sentence stands in for Gödel numeral of that sentence).

Proof. See [4], pp. 167.

Second Condition. T contains an arithmetical predicate $I(x, y)$ meaning that the sentence with Gödel number x is derivable from the sentence with Gödel number y, the Gödel numbering being given in advance. Note that in T we can formalize the proof predicate $Bew(x)$, normally defined as $\exists y.Proof_T(y, x)$, which satisfies the rule (for the proof see [2], pp. 43):

$$\vdash \alpha \quad \Rightarrow \vdash Bew(\overline{\alpha})$$

From this fact it follows that if T satisfies the deduction theorem, then $I(\overline{\alpha}, \overline{\beta})$ is expressible as $Bew(\overline{\alpha \to \beta})$.

Third Condition. T contains a unary predicate $K(x)$ meaning that the truth of the sentence with Gödel number x is known and satisfying, for every sentence α and β of $\mathcal{L}$(T):

$K(\overline{\alpha}) \to \alpha$, the reflection principle (RP),

$K(\overline{K(\overline{\alpha}) \to \alpha})$, assumption number two (AT),

$K(\overline{\alpha}) \wedge I(\overline{\alpha}, \overline{\beta}) \to K(\overline{\beta})$, epistemic closure (EC).

Note that if T satisfies the deduction theorem and it contains the rule **R.K.**:

$$\vdash \alpha \quad \Rightarrow \vdash K(\overline{\alpha})$$

then the principle of epistemic closure can be expressed in the following alternative form:

$$K(\overline{\alpha \to \beta}) \to (K(\overline{\alpha}) \to K(\overline{\beta})), \text{ epistemic closure } (EC^1).$$

The Knower Paradox. If the theory T satisfies these three conditions, then T is inconsistent.

Proof. By the fixed point lemma, there is a sentence $\neg p$ of $\mathcal{L}(\mathrm{T})$ such that:

(1) $T \vdash \neg p \leftrightarrow \neg K(\overline{\neg p})$

(2) $T \vdash \neg\neg p \leftrightarrow \neg\neg K(\overline{\neg p})$, by contraposition from (1),

(3) $T \vdash p \leftrightarrow K(\overline{\neg p})$, by logic from (2),

(4) $(RP) \vdash_T K(\overline{\neg p}) \to \neg p$,

(5) $(RP) \vdash_T p \to \neg p$, by transitivity from (3) and (4),

(6) $(RP) \vdash_T \neg p$, by logic from (5),

(7) $T \vdash I(\overline{RP}, \overline{\neg p})$, from what precedes,

(8) $(AT) \vdash_T K(\overline{RP})$,

(9) $(EC) \vdash_T K(\overline{RP}) \wedge I(\overline{RP}, \overline{\neg p}) \to K(\overline{\neg p})$,

(10) $(EC), (AT) \vdash_T K(\overline{\neg p})$, by M. P. form (7), (8), (9),

(11) $(EC), (AT) \vdash_T p$, by logic from (10) and (3).

(12) $(EC), (AT), (RP) \vdash_T p \wedge \neg p$, combining (11) with (6).

(q.e.d.)

Even if the Knower Paradox in its original formulation was presented as a Paradox affecting the framework of first-order arithmetic, it can be easily reproduced in the framework of modal logic.

Let L be a *normal* system of modal logic in the modal language $\mathcal{L}$: so L contains all tautologies, all instances of the distribution axiom schemata **K** $[\Box(\alpha \to \beta) \to (\Box\alpha \to \Box\beta)]$, it is closed under modus ponens and the rule of necessitation **R.N.** [$\vdash \alpha \Rightarrow \vdash \Box\alpha$] and it satisfies the deduction theorem. Let L^* a modal system in the modal language $\mathcal{L}^*$ that satisfies the following two conditions.

First condition. L^* extends L by proving a fixed point lemma.

The modal fixed point lemma. For every $\mathcal{L}$ - formula $\alpha(p, q_1, ..., q_n)$, in which every occurrence of p is in the scope of some occurrence of $\Box$, there exists a sentence $\beta(q_1, ..., q_n)$, where β contains no sentence letter except $q_1, ..., q_n$ and p is distinct from the q_i's, such that:

$$L^* \vdash \beta(q_1, ..., q_n) \leftrightarrow \alpha(\beta(q_1, ..., q_n), q_1, ..., q_n).$$

Proof. See, [2], pp. 141-145.

Second Condition. L^* contains the axiom schemata, **T**: $\Box\alpha \rightarrow \alpha$

Let us make three observations. Firstly, **T** is quite obviously the modal counterpart of the reflection principle; it suffices to substitute in the reflection principle K by $\Box$, to obtain **T**:

$$K(\overline{\alpha}) \rightarrow \alpha \quad \rightsquigarrow \quad \Box\alpha \rightarrow \alpha$$

Secondly, if L^* contains **T** and the rule of necessitation, then L^* also contains the axiom schemata **U**: $\Box(\Box\alpha \rightarrow \alpha)$, which is the modal counterpart of the assumption number two. Again it suffices to substitute in the assumption number two K by $\Box$, to obtain **U**:

$$K(\overline{K(\overline{\alpha}) \rightarrow \alpha}) \quad \rightsquigarrow \quad \Box(\Box\alpha \rightarrow \alpha)$$

Thirdly, as L^* contains the distribution axiom schemata and the rule of necessitation, then L^* contains the modal counterpart of the principle of epistemic closure in his alternative form (EC^1). It suffices to substitute in the rule **R.K.** and in EC^1 K by $\Box$, to obtain the rule **R.N.** and the distribution axiom schemata **K**:

$$\vdash \alpha \quad \Rightarrow \quad \vdash K(\overline{\alpha}) \quad \rightsquigarrow \quad \vdash \alpha \quad \Rightarrow \quad \vdash \Box\alpha$$

$$K(\overline{\alpha \to \beta}) \to (K(\overline{\alpha}) \to K(\overline{\beta})) \quad \rightsquigarrow \quad \Box(\alpha \to \beta) \to (\Box\alpha \to \Box\beta)$$

The Knower Paradox in the Framework of Modal Logic. If L^* satisfies these two conditions, then L^* is inconsistent.

Proof. This result follows immediately from the inconsistency result just derived in the first-order arithmetic theory.

2. *Two important points common to the three solutions*

The aim of this article is to present three solutions to the Knower Paradox. These solutions agree on the following two points. Firstly, they agree on the source of the Paradox. In order to solve the Paradox, we must find an actual mistake in the argument. This can be: a fallacious assumption, an invalid inference or an inconsistent theory (or system). Given that the theory T (or the system L^*) is supposed to be consistent, or at least possibly consistent, and the proof is trivially valid, then one of the three assumptions, namely RP, AT and EC (or their modal counterparts, namely **T**, **U** and **K**) must be at fault and consequently, to avoid the Paradox, we have to renounce to one of them. The three solutions agree that the assumption number two is the source of the contradiction and I want to give three reasons which support this claim.

Firstly, the other two assumptions are relatively difficult to doubt, if we attentively analyze them. A recent and really interesting result of C.B. Cross [6] and [7] tells us that we can prove the Knower Paradox without an explicit epistemic closure. If we can derive the Knower Paradox without the epistemic closure, then we can draw two conclusions: the first is that the principle of epistemic closure can't be the source of the Paradox, the second is that the culprit of the contradiction has to be the reflection principle or the assumption number two. Let us then analyze these two assumptions. The reflection principle is the most natural and intuitive one. More, it's supported by our standard conception of knowledge, the one which defines knowledge as ⟪justified true belief⟫, which has been firstly introduced by Plato in the Theaetetus (see, e.g., [5]). As we do want to preserve this conception of knowledge, we can't refuse the reflection principle. So, it is useless to reject the principle of epistemic closure because the contradiction can be proved without it; it is implausible to reject the reflection principle without rejecting knowledge itself; hence the only assumption we can reject, to avoid the Knower Paradox, is the assumption number two.

Furthermore there are two arguments that suggest that the assumption number two is the source of the Paradox. The first argument relies on the analogy with provability and the Gödel's second theorem. Let's consider the three assumptions which are satisfied in the formal system T: the principle of reflection, the assumption number two and the principle of epistemic closure. If we substitute in every assumption the knowledge predicate K by the provability predicate Bew, we obtain:

$$Bew(\overline{\alpha}) \to \alpha,$$

$$Bew(\overline{Bew(\overline{\alpha}) \to \alpha}),$$

$$Bew(\overline{\alpha}) \wedge Bew(\overline{\alpha}, \overline{\beta}) \to Bew(\overline{\beta})$$

But $Bew(\overline{\neg Bew(\overline{\bot})})$, which contradicts the Gödel's second theorem, is just an instance of $Bew(\overline{Bew(\overline{\alpha}) \to \alpha})$. Therefore, quoting Anderson ([1], p. 350), ⟪the thing that's false for Q is the analogue of the assumption number two. Shouldn't we do the analogous thing for our case? ⟫

The second argument is based on the form of the three assumptions. The reflection principle and the epistemic closure describe some conditions and features of knowledge: the former says that *if* something is known, then it is so; the latter says that *if* β is derivable from α and α is known, then β is known. As we can trivially see, they do not say anything about what we really do know, whereas the assumption number two affirms knowledge of something (the reflection principle). This difference of shape could be a good sign for believing that the assumption we have to rule out is the second one. As Anderson says ([1], p. 350), 《None of this is conclusive. Probably there's a more convincing argument here somewhere, but I can't find it》.

Having presented the first point common to the three solutions, let us now pass to introduce the second one. Such point relates to the notion of knowledge. Indeed each of the three solutions to the Knower Paradox treats knowledge as if it were intimately connected with the notion of proof. They thus seem to deal with a more limited and abstract notion of knowledge than that which is generally accepted; so they appear as solutions to the the Paradox of a Mathematical Knower, and not the Paradox of an Ordinary Knower. As we will see in the fourth section the main difference between Anderson's solution and the one offered by the modal logic of provability is based on this point: while Anderson appears to confuse two different notions of proof by which we obtain knowledge, the logic of provability uses a unique and clear notion of proof.

3. *The first solution: Anderson's solution*

Anderson's solution, published in 1983 in the Journal of Philosophy, is, as we will see, a hierarchical solution expressed in the frame of first order arithmetic.

In order to be able to present Anderson's solution, we have, first of all, to consider a language Q^* extending the language of Robinson Arithmetic with additional non-logical constants. Then we have to isolate K_0, a recursively enumerable set of Q^*-formulas, containing things expressible in Q^* and known by a certain person James. Now we can construct a formal system, call it Q^1, whose theorems are the sentences derivable from the axioms of Robinson Arithmetic together with the sentences of K_0. Is Q^1 consistent? Yes it is: the axioms of Q (Robinson

Arithmetic) are true in the standard model, the things contained in K_0 are known by James and hence they are true, the rules of inferences preserve truth, so every theorem of Q^1 is true. But in Q^1, as it contains Robinson Arithmetic, we can formulate the Gödel sentence $G \leftrightarrow \neg Bew(\overline{G})$. Is G provable? No, if it were, $\neg Bew(\overline{G})$ would be provable, and so a falsity would be provable contrary to the truth of every theorem of Q^1. Does James know G? No, G is not in K_0, else it would be provable in Q^1. But *I* have just given a proof that G is not provable, hence I know G, or G is contained in the set K_1 of the things that I know. If I were James, Anderson argues, it would seem that we have to distinguish different levels of knowledge: what I know at stage K_0, and what I know at stage K_1, after reflecting on Gödel's theorem.

The conclusion is that we have to distinguish levels of knowledge – $K_0, K_1, K_2, \ldots$ – and *this idea of levels of knowledge has to be applied to the assumption number two*, considered to be the source of the Paradox. Anderson formalizes such a conclusion with the following hierarchy of languages:

Languages. Let $\mathcal{L} = \mathcal{L}(Q)$ be the language of Robinson Arithmetic, we have:

$$\mathcal{L}_0 = \mathcal{L} \cup \{K_0, I_0\}^1$$

$$\mathcal{L}_{n+1} = \mathcal{L}_n \cup \{K_{n+1}, I_{n+1}\}$$

$$\mathcal{L}_\omega = \bigcup_{n\in\omega} \mathcal{L}_n$$

So starting with the language of Robinson Arithmetic, Anderson constructs a hierarchy of theories $(T_n)_{n\in\omega}$ satisfying:

[1]Note that Anderson adds inferences predicates I_n to his hierarchy, but he does it with minimal argument; this point is not, however, central to my discussion.

$$T_0 = Q \cup \begin{cases} K_0(\overline{\alpha}) \rightarrow \alpha \\ \\ K_0(\overline{\alpha}) \wedge I_0(\overline{\alpha}, \overline{\beta}) \rightarrow K_0(\overline{\beta}) \end{cases}$$

$$T_{n+1} = T_n \cup \begin{cases} K_{n+1}(\overline{\alpha}) \rightarrow \alpha \\ \\ K_{n+1}(\overline{K_n(\overline{\alpha}) \rightarrow \alpha}) \\ \\ K_{n+1}(\overline{\alpha}) \wedge I_{n+1}(\overline{\alpha}, \overline{\beta}) \rightarrow K_{n+1}(\overline{\beta}) \end{cases}$$

where α and $\beta \in L_\omega$. As we can trivially see, the assumption number two has now the form: $K_{n+1}(\overline{K_n(\overline{\alpha}) \rightarrow \alpha})$ which is always true, while $K_n(\overline{K_n(\overline{\alpha}) \rightarrow \alpha})$ is, in some cases, false. In this way, Anderson avoids the Paradox of The Knower.

4. *My criticism to Anderson's solution*

My criticism of Anderson's solution is based on two arguments: a principal and specific argument and a second general argument which is inspired by S. Kripke's article of 1975, [8].

The main argument can be succinctly expressed by the following sentence: I argue that, in his solution, which relies on a certain relationship between knowledge and proof (see, section 2), Anderson confuses two different notions of proof: a *syntactic* one and an *absolute* one. So in order to be able to explain my criticism, firstly I have to introduce these two notions of proof. The syntactic notion of proof is well known and it is linked with the notion of formal system that is a set of sentences, called the axioms of the system, together with a set of relations on sentences, called the rules of inference of the system. A proof of a sentence in a system, or a syntactic proof, is a finite sequence of sentences, each of which either is an axiom of the system, or is immediately deducible from the earlier ones by one of the rules of inferences. One of the possible interpretations of Gödel famous theorem deals with this notion of proof: if a formal system satisfies certain conditions, there exists a formula p such that neither p nor $\neg p$ is formally provable in that system. But in Myhill's opinion (see [12]) there is another quite interesting and problematic interpretation of Gödel's theorem which is expressible in the following way: for any formal system S, adequate for the arithmetic of natural numbers, there are *correct inferences* which cannot be formalized in S. So the question is: what does Myhill mean when he uses the notion of correct inferences? Let's try to explain such a concept. When a person uses some methods of proof, he is committed to the validity of his methods of proof, otherwise he would be contradicting himself. Hence if someone uses the means furnished by the system of Peano Arithmetic (or Robinson Arithmetic) for proving arithmetical statements, and is able to make explicit those methods, « he is *ipso facto* committed to the perfectly definite proposition that the use of his methods cannot lead to a false arithmetical statement» ([12], p. 462). So let us consider a person who uses the means furnished by the system of Peano Arithmetic for proving arithmetical statements and who succeeds in making these means explicit. This person will be able to reason as follows: the axioms of Peano Arithmetic are true (he is committed to this, because he is committed to the validity of his proof methods), and the rules of inferences preserve truth (*idem*), therefore every theorem of Peano Arithmetic is true. Hence a false statement, such as $0 = 1$, is not a theorem, so (the arithmetization of) the statement that says that $0 = 1$ is not a theorem is true. However, such reasoning cannot be formalized in the system of Peano Arithmetic. It is thus not a syntactic proof. It can only be thought of as a proof in a different sense, which we will call, following Myhill's definition, *absolute,* and in this sense of proof the undecidable sentence p is demonstrable.

Since I have clarified the meaning of the syntactic and the absolute notion of proof, I am now able to explain my criticism to Anderson's solution. At the beginning of his discussion, Anderson adopts a notion of knowledge based on a formal notion of proof: the set K_0 of the things known by James is recursively enumerable and contains Q, Robinson Arithmetic (as we will see in a moment, K_0 has to contain Q, otherwise James could never be me, as Anderson wants). Hence James knows by a syntactic notion of proof, and so, for what we have just said, he doesn't know the Gödel sentence G: nobody, who knows by this sense of knowledge, could never know the Gödel sentence G. So how could James know G? He has to use the proof I have utilized for knowing G, but that kind of proof is not at all a syntactic one: on the contrary is the sort of proof we have defined as an absolute notion of proof. So Anderson'solution invites two important questions. The first one is: how can Anderson define, at the beginning of his argument, James's knowledge as knowledge based on a syntactic notion of proof, and then adopt a definition of James's knowledge based on an absolute notion of proof ? The second one is: if Anderson had to give a definition of the concept of knowledge, on which of the two notions of proof would depend ? The trivial answer is that there is no reason for changing the notion of proof on which James's knowledge is based, and that, therefore, Anderson cannot give a definite and clear notion of knowledge.

This was the first argument against Anderson's solution. Let us now pass to the second and more general one. The second argument I want to propose is based, as I have already said, on the analogy with Kripke's criticism of the Tarski's idea of introducing a hierarchy of languages each with its own predicate of truth for the preceding language. In Kripke's opinion, Tarski's theory goes against our intuitive usage of the predicate "being true": in the natural language there is *only one* truth predicate and not many. In the same way, we can argue against Anderson's solution that in the natural language we have only one knowledge predicate. So the idea of introducing many levels of knowledge is not only based on a incorrect argument, but is also quite counterintuitive.

5. *The second solution: modal logic of provability.*

The second solution is based on the idea of giving an epistemic interpretation of the modal logic of provability. Let us, first, explain what is the modal logic of provability.

The modal logic of provability, developed only in the early 70's and based on the two systems GL and GL^*, is a logic that tries to answer the following question: can we capture the features of the provability predicate $Bew(x)$ in the framework of modal logic? The positive answer to this question, given by R. Solovay in 1976 [15], comes in the form of an equivalence between the system PA (Peano Arithmetic) and the two systems GL and GL^*. In order to explain such an answer in a mathematically precise way, we have to present the axiomatization of GL and GL^* and introduce the notions of realization and translation. GL is a normal system of modal logic characterized by the two axiom schemata:

$$\mathbf{4} : \Box\alpha \to \Box\Box\alpha$$

$$\mathbf{G} : \Box(\Box\alpha \to \alpha) \to \Box\alpha$$

On the other hand, the axioms of GL^* are the theorems of GL and all sentences of the form $\Box\alpha \to \alpha$ ($\mathbf{T}$), and the sole rule of inference is modus ponens. GL^* is not a normal system of modal logic as it is not closed under the rule of necessitation.

Let us now introduce the two notions of realization and translation.

A *realization* is a function that assigns to each sentence letter a sentence of the language of Peano Arithmetic.

A *translation* α^ϕ of a modal sentence α under a realization ϕ is defined inductively:

$$\perp^\phi = \perp$$

$$p^\phi = \phi(p)$$

$$(\alpha \rightarrow \beta)^{\phi} = (\alpha^{\phi} \rightarrow \beta^{\phi})$$

$$(\Box\alpha)^{\phi} = Bew(\overline{\alpha^{\phi}})$$

These definitions allow us to express Solovay's result:

$$GL \vdash \alpha \quad iff, \quad for \quad every \quad \phi, \quad PA \vdash \alpha^{\phi}$$

$$GL^{*} \vdash \alpha \quad iff, \quad for \quad every \quad \phi, \quad \models_{\mathcal{N}} \alpha^{\phi}$$

The central idea of the second solution is to use the system GL^* (which is a consistent system of modal logic even if we add the fixed point lemma) *to resolve the Paradox of the Knower.* In order to develop such an idea, we have to give an epistemic interpretation of GL^*. Let us then consider GL^*, not as a modal logic system, but as an *idealized mathematician knower* who has a fairly complete knowledge of mathematics. The theorems of GL^*, axiomatized by the principles we have given on page 9, are exactly what our mathematician knows. Hence the set of formulas he knows:

> contains only, but not all, arithmetical truths (modulo ϕ),
>
> is closed under modus ponens.

Finally his knowledge has the following features:

- **K** : it is closed under modus ponens,
- **T** : it satisfies the reflection principle,
- **4** : if he knows something then he knows that he knows it,
- **G** : ≪Our knower couldn't be more modest about his knowledge: he never knows to be sound with regard to an unknown sentence≫ (see [14], pp. 350-351).

Remark that the modal logic of provability substitute the problematic assumption number two, whose modal counterpart we have designed with the axiom schemata **U** (see, p.3), with the new axiom **G**. **G** has a form which is suitable for the observations we have made before: it does not affirm knowledge of something, but it just describes a feature of knowledge as the reflection principle and the principle of epistemic closure do. *So thanks to an idealized and abstract notion of knowledge and to the consistent and decidable system GL^*, we menage to avoid the Paradox of the (Mathematical) Knower.*

6. *The third solution: Egré's solution*

We arrive finally at the third and last solution, which tries to connect the first two. Indeed P.Egré in his paper [8] arrives at the conclusion of the section 5 above and then, instead of stopping *as he should*, argues that there is a link between the modal logic of provability and the first system of Anderson's hierarchy. What sort of connection does he propose? First of all P. Egré affirms that GL^* *corresponds* to PA^+, where PA^+ contains as axioms the theorems of PA and all sentences of the form $Bew(\overline{\alpha}) \rightarrow \alpha$, and it is closed under modus ponens. Then he says that:

$$PA^+ = T_0$$

where T_0 is the first system of Anderson's hierarchy (see, p.6).

Hence in Egré's opinion, Anderson's solution and the solution offered by the modal logic of provability are not so different: ≪ Anderson's hierarchy is a generalization to all the finite degrees of the separation of axiom schemata reflected in Solovay's system GL^*≫ ([8], p. 45). However this conclusion is incorrect. For a start, it runs into the following technical problems. Namely, we will show, contra Egré, that:

$$PA^+ \neq T_0$$

In order to prove such a formula, we have to analyze the assertion "GL^* *corresponds* to PA^+" which is quite unclear since there is no notion of realization and translation, nor theorem as the Solovay's one. So there are two possible interpretations:

$$(1) \quad (\Box\alpha)^\phi = K(\overline{\alpha^\phi})$$

Then $PA^+ \neq T_0$, for two reasons. The first one is that PA^+ contains the translation of the axiom schemata **G** and **4**, whereas T_0 does not. The second reason is that PA^+ contains the translation of the distribution axiom schemata **K** which is not equivalent to the principle of epistemic closure contained in T_0. In fact **K** is only equivalent to $K(\overline{\alpha}) \wedge I(\overline{\alpha}, \overline{\beta}) \rightarrow K(\overline{\beta})$, in the presence of the translation of the rule of necessitation, that is not, as we have already said, a rule of PA^+. The other possible interpretation is:

$$(2) \quad (\Box\alpha)^\phi = Bew(\overline{\alpha^\phi})$$

Then $PA^+ \neq T_0$, for two reasons. The first one is that, as before, PA^+ does not contain a 'real epistemic closure' whereas T^0 does. The second, and maybe more important reason, relies on the tension between the following two facts. On the one hand, as we have already said in section 1, in any arithmetical theory where we can define the proof predicate $Bew(x)$ as $\exists y.Proof_T(y, x)$, $Bew(x)$ satisfies the rule $\vdash \alpha \Rightarrow \vdash Bew(\overline{\alpha})$. On the other hand, we want the arithmetical theory PA^+ to contain a proof predicate $Bew(x)$, defined as $\exists y.Proof_T(y, x)$, which *does not* satisfy the rule $\vdash \alpha \Rightarrow \vdash Bew(\overline{\alpha})$, since PA^+ is the arithmetical counterpart of GL^*. It is clear that these two facts are incompatible (it is no accident that the Solovay theorem hold between GL^* and the standard model, and not, as for GL, between GL^* and a formal system).

These technical weaknesses of Egré'solution may not be unrelated to the philosophical points underlined in the preceding discussion. In fact the modal logic of provability and Anderson's solution employ different means: while Anderson uses, in an illegitimate way, two notions of proof (the absolute one and the syntactic one) to define knowledge, in the modal logic of provability it's clear that to know means to prove in a formal system.

References

[1] C. A. Anderson, *The Paradox of the Knower*, 《Journal of Philosophy》, 80, 1983, pp. 338-355.

[2] G. Boolos, *The Unprovability of Consistency*, Cambridge University Press, Cambridge 1979.

[3] G. Boolos, *The Logic of Provability,* Cambridge University Press, Cambridge 1993.

[4] G. Boolos and R. Jeffrey, *Computability and Logic,* Cambridge University Press, New York 1974.

[5] L. Campbell, *Theaetetus of Plato,* Oxford University Press, Oxford, 1883.

[6] C.B. Cross, *The Paradox of the Knower without Epistemic Closure*, 《Mind》, 110, 2001, pp. 319-333.

[7] C.B. Cross, *More on the Paradox of the Knower without Epistemic Closure*, 《Mind》, 113, 2004, pp. 109-114.

[8] P. Egré, *The Knower Paradox in the Light of Provability Interpretations of Modal Logic,* 《 Journal of Logic, Language and Information》, 14, 2005, pp. 13-48.

[9] D. Kaplan and R. Montague, *A Paradox Regained*, 《 Notre Dame Journal of Formal Logic》, 1, 1960, pp. 79-90.

[10] S. Kripke, *An Outline of a Theory of Truth,* 《Journal of Philosophy》, 72, 1975, pp. 690-716.

[11] R. Montague, *Syntactical Treatments of Modality, with Corollaries on Reflection Principle and Finite Axiomatizability,* 《Acta Philosophica Fennica》, 16, 1963, pp. 153-167.

[12] J. Myhill, *Some Remarks on the Notion of Proof,* 《Journal of Philosophy》, 14, 1960, pp. 461-471.

[13] R.M. Smullyan, *Logicians who Reason about Themselves,* in *Reasoning About Knowledge, Proceedings of the TARK Conference,* Morgan Kaufman, San Mateo, CA 1986, pp. 341-352.

[14] R.M. Smullyan, *Gödel's Incompleteness Theorems,* Oxford Logic Guides 19, Oxford University Press, New York 1992.

[15] R. Solovay, *Provability Interpretations of Modal Logic*, 《Israel Journal of Mathematics》, 25, 1976, pp. 287-304.

[16] R. Sorensen, *Formal Problems about Knowledge,* in *The Oxford Handbook of Epistemology*, Oxford University Press, New York 2002, pp. 539-568.

[17] T. Tymoczko, *An Unsolved Puzzle about Knowledge,* 《Philosophical Quarterly》, 34, 1984, pp. 437-458.