Minds & Machines (2012) 22:1–24
DOI 10.1007/s11023-011-9260-0

On the Behavior of True and False

Stefan Wintein

Received: 10 March 2011 / Accepted: 25 September 2011 / Published online: 13 October 2011

Abstract Uzquiano (Analysis 70:39–44, 2010) showed that the Hardest Logic Puzzle Ever (*HLPE*) [in its amended form due to Rabern and Rabern (Analysis 68:105–112, 2008)] has a solution in only two questions. Uzquiano concludes his paper by noting that his solution strategy naturally suggests a harder variation of the puzzle which, as he remarks, he does not know how to solve in two questions. Wheeler and Barahona (J Philos Logic, to appear, 2011) formulated a three question solution to Uzquiano's puzzle and gave an information theoretic argument to establish that a two question solution for Uzquiano's puzzle does not exist. However, their argument crucially relies on a certain conception of what it means to answer *self-referential* yes–no questions *truly* and *falsely*. We propose an alternative such conception which, as we show, allows one to solve Uzquiano's puzzle in two questions. The solution strategy adopted suggests an even harder variation of Uzquiano's puzzle which, as we will show, can also be solved in two questions. Just as all previous solutions to versions of *HLPE*, our solution is presented informally. The second part of the paper investigates the prospects of formally representing solutions to *HLPE* by exploiting theories of truth.

Keywords Hardest logic puzzle ever · Self-reference · Truth

Introduction

Recall Boolos' formulation of the *Hardest Logic Puzzle Ever* (*HLPE*):

> The Puzzle: Three gods A, B and C are called, in some order, True, False, and Random. True always speaks truly, False always speaks falsely, but whether

S. Wintein (✉)
Department Wijsbegeerte and TiLPS, Universiteit van Tilburg, Warandelaan 2, Gebouw D136, 5000 LE Tilburg, The Netherlands
e-mail: s.wintein@uvt.nl

> Random speaks truly or falsely is a completely *random* matter. Your task is to determine the identities of A, B, and C by asking three yes–no questions; each question must be put to exactly one god. The gods understand English, but will answer all questions in their own language, in which the words for 'yes' and 'no' are 'da' and 'ja' in some order. *You do not know which word means which*. Before I present the somewhat lengthy solution, let me give answers to certain questions about the puzzle that occasionally arise:
>
> - (B1) It could be that some god gets asked more than one question (and hence that some god is not asked any question at all).
> - (B2) What the second question is, and to which god it is put, may depend on the answer to the first question (and of course similarly for the third question).
> - (B3) Whether Random speaks truly or not should be thought of as depending on the flip of a coin hidden in his brain: if the coin comes down heads, he speaks truly, if tails, falsely.
> - (B4) Random will answer 'da' or 'ja' when asked any yes–no question.
>
> Boolos (1996), p. 62

Rabern and Rabern (2008) point out the need to distinguish *HLPE* as literally formulated by Boolos from a version of *HLPE* which is closely related to it and, as pointed out by Rabern and Rabern, is more properly called 'the hardest logic puzzle ever'. The distinction between the puzzle as formulated by Boolos—which we call $HLPE_{sem}$, for *semantic HLPE*—and the amended puzzle—which we call $HLPE_{syn}$, for *syntactic HLPE*—only concerns the way in which Random reacts to questions. Suppose that we address a question to Random. Depending on the version of *HLPE* under consideration, he reacts as follows:

- $HLPE_{sem}$: Random flips a coin and then, depending on the outcome of the coin-flip, answers the question either truly or falsely.
- $HLPE_{syn}$: Random flips a coin and then, depending on the outcome of the coin-flip, answers the question with either 'da' or 'ja'.

Rabern and Rabern show that $HLPE_{sem}$ allows for a solution (in three questions) which is so simple that it almost trivializes the puzzle. Previous commentators (such as Boolos 1996 and Roberts 2001) did not realize the possibility of such a simple solution and Rabern and Rabern plausibly suggest that this is due to the fact that these commentators implicitly assumed that Random worked along the lines of $HLPE_{syn}$. Accordingly, we may regard $HLPE_{syn}$ as a corrected version of $HLPE_{sem}$ which is more properly called 'the hardest logic puzzle ever'.

Besides pointing out the distinction between $HLPE_{sem}$ and $HLPE_{syn}$, Rabern and Rabern come up with a solution to $HLPE_{sem}$ which exploits only two (!) questions. To realize their solution, Rabern and Rabern ask the gods *self-referential questions*, which, as they observe, is not prohibited by Boolos' guidelines. However, their solution does not carry over to $HLPE_{syn}$ and so the question arises whether $HLPE_{syn}$ allows for a two-question solution as well.

Uzquiano (2010) shows that $HLPE_{syn}$ has a two-question solution.[1] His solution strategy is inspired by Rabern and Rabern's observation that, given their nature, True and False cannot answer all yes–no questions with 'da' and 'ja'. In Uzquiano's framework, True and False are said to *remain silent* on questions that they cannot answer with 'da' or 'ja'. Assuming for simplicity that the gods understand and answer in English, an example of a question on which True must remain silent is given by λ.

λ: Is it the case that: your answer to λ is 'no' ?

In answering λ with either 'yes' or 'no', True can be accused of lying and so True cannot answer λ "in accordance with his nature". Accordingly, True will remain silent when asked λ. This illustrates that in $HLPE_{syn}$, True and False are thought of as having three reactions to questions; besides answering with 'da' and 'ja' they may also remain silent. However, $HLPE_{syn}$ models Random as a random variable over only two of these reactions: answering with 'da' or answering with 'ja'. As Uzquiano observes, a more natural way to model Random in $HLPE_{syn}$ then, is as a *ternary* random variable, the outcome of which determines whether Random answers 'da', 'ja' or remains silent. $HLPE^2_{syn}$ and $HLPE^3_{syn}$ will be used to denote the original version and Uzquiano's version of $HLPE_{syn}$ respectively. Uzquiano solves $HLPE^2_{syn}$ in two questions, but with respect to $HLPE^3_{syn}$, he remarks that: 'I, for one, do not know how to solve this puzzle in two questions.'

Wheeler and Barahona (2011) give a three-question solution to $HLPE^3_{syn}$ and give an information theoretic argument which establishes that $HLPE^3_{syn}$ cannot be solved in less than three questions. Although their argument is certainly correct, it crucially relies on the assumption that there are three distinct ways in which the gods answer yes–no questions. But now consider what happens if we ask the following question to True.

τ: Is it the case that: your answer to τ is 'yes' ?

Indeed, just as λ is (when asked to True) an interrogative version of the *Liar*, so τ is (when asked to True) an interrogative version of the *Truthteller*. And just as the Truthteller may be valuated as either true or false, so True can answer τ with either 'yes' or 'no'. However, doing so is, in both cases, completely *arbitrary*. Questions like τ do not have a role to play in previous solutions to *HLPE*, and none of the mentioned papers discusses how True answers such questions. In this paper however, questions like τ will have a crucial role to play: they give rise to a *fourth answer*. Exploiting a four-valued answering repertoire, we will show how to solve $HLPE^3_{syn}$ in two questions.

[1] Actually, Uzquiano distinguishes two versions of $HLPE_{syn}$ and gives two-question solutions for both versions. The versions differ with respect to the abilities of True and False to predict the answers of Random. The first version assumes that True and False cannot predict Random's answers (which seems reasonable given that Random answers *randomly*), while the second version assume that True and False can predict Random's answers (which seems reasonable as True and False are *omniscient*). Uzquiano's solution to the second version is also a solution to the first version but not vice versa. For our purposes, the distinction does not matter: we give a solution that works for both versions.

Our alternative account of how True and False answer yes–no questions makes the arbitrariness of answering τ with either 'yes' or 'no' explicit. According to our account, True gives the following answers to λ and τ:

λ can *neither* be answered with 'yes' nor with 'no'
τ can be answered *both* with 'yes' and 'no'

There is a clear intuitive sense in which answering λ and τ as such is speaking truly. False will answer the mentioned questions as follows.

λ can *both* be answered with 'yes' and with 'no'
τ can be answered *neither* with 'yes' nor with 'no'

Again, there is a clear intuitive sense in which answering λ and τ as such is speaking falsely. A possible justification of working with a four-valued (in contrast to a three-valued) answering repertoire is that the four (linguistic) answers allow us, even in the presence of self-reference, to respect Boolos' instructions, which state that 'True *always speaks* truly' and 'False *always speaks* falsely'. On the other hand, we can also interpret our two non-standard answers in non-linguistic terms, along the following lines: on questions like λ, the algorithm which describes True's behavior yields no solutions, while on questions like τ it yields two solutions. In such cases, True does not answer with 'yes' or 'no', but its two (non-linguistic) answers reflect, respectively, the lack and abundance of solutions. Although we will work with the linguistic version of our non-standard answers, i.e., with 'both' and 'neither', we will return to the two distinct justifications of a four-valued answering repertoire (cf. section "Critical Remarks on Formalizations").

An answering repertoire of four answers naturally suggests an "even harder" variation of the hardest logic puzzle ever, $HLPE^4_{syn}$, in which Random is modeled as a four-valued random variable over the possible answers. As we will see, $HLPE^4_{syn}$ can also be solved in two questions. In fact, we will only show how to solve $HLPE^4_{syn}$ in two questions as our solution to $HLPE^4_{syn}$ is easily seen to solve $HLPE^3_{syn}$ as well.[2]

All previous solutions to *HLPE* are presented (informally) in natural language and our solution to $HLPE^4_{syn}$, as presented in section "Solving the Puzzles", is no exception. However, given the nature of the gods True and False, one would expect that solutions to *HLPE* allow for a formal representation that is based on a (formal) theory of truth. In section "Formalizations via Theories of Truth", we explore the prospects of such a formal representation, exploiting (Kripkean) fixed point theories of truth. We will see that, using a restricted formal language, the previous solutions to *HLPE* as well as the solution put forward in section "Solving the Puzzles", can be given a formal representation. The formal representations are illuminative as they clearly lay bare the differences between the previous solutions to *HLPE* and the present one. Although our formalization allows us to represent the (informal) solutions to *HLPE*, nevertheless there are some reasons for not being completely satisfied with it, as will be explained in section "Formalizations via Theories of Truth". Section "Formalizations via Theories of Truth" concludes by discussing the

[2] Note that $HLPE^3_{syn}$ and $HLPE^4_{syn}$ (deliberately) violate Boolos' instruction (*B*4).

Table 1 Reactions of True and False

Q(uestion)	Y/N	$\mathcal{V}(Q)$	✔/X	True	False
sw: snow is white	$Y(sw)$	True	✔	Yes	No
	$N(sw)$	True	X		
sb: snow is black	$Y(sb)$	False	X	No	Yes
	$N(sb)$	False	✔		
λ: $N(\lambda)$	$Y(\lambda)$	False	X	Neither	Both
	$N(\lambda)$	True	X		
τ: $Y(\tau)$	$Y(\tau)$	True	✔	Both	Neither
	$N(\tau)$	False	✔		

information theoretic argument of Wheeler and Barahona (2011), which establishes that (given a three-valued answering repertoire) $HLPE^3_{syn}$ cannot be solved in less than three questions. Section "Concluding Remarks" concludes the paper.

Solving the Puzzles

Gods Who Answer with 'Yes' and 'No'

In this section, we solve $HLPE^4_{syn}$ under the assumption that the gods speak English: they use 'yes' and 'no' to answer positively and negatively respectively. In the next section we give up this simplifying assumption and show how to solve $HLPE^4_{syn}$ itself, in which the gods answer with 'da' and 'ja'.

We use the following abbreviations. A, B and C will be used as in Boolos' guidelines and T, F and R will be used to denote True, False and Random respectively. With x an arbitrary question, $N(x)$ reads as 'your answer to x is 'no'', while $Y(x)$ reads as 'your answer to x is 'yes''.[3] Before we state our solution to (the English version of) $HLPE^4_{syn}$, we first briefly comment on the algorithm that gives rise to the answers of True and False. First, True and False calculate how their yes/no answers to a question Q influence the truth-value[4] of Q, in light of which they judge these yes/no answers to be correct (✔) or incorrect (X). Exploiting the correctness/incorrectness of their yes–no answers with respect to Q, they then determine which of the four possible answers ('yes', 'no', 'both', 'neither') they give to Q. The process is illustrated by the Table 1.

Clearly, the yes–no answers of the gods to *sw* do not influence its truth-value (which is true). Accordingly, answering *sw* with 'yes' is correct while answering with 'no' is incorrect. Accordingly, True will answer *sw* with 'yes' while False answers it with 'no'. The yes–no answers of the gods to λ do influence its truth-value. As illustrated by Table 1, answering λ with either 'yes' or 'no' is incorrect.

[3] We could use two place answering predicates and remove the indexical "your". However, as this results in a less streamlined presentation, we chose not to do so.

[4] We treat yes–no questions on par with their associated yes–no statements. That is sloppy, but also very convenient.

Table 2 Reactions of True and False on α_1

World	*Y*/*N*	$\mathcal{V}(\alpha_1)$	✔/X	True	False
$A = R$	$Y(\alpha_1)$	False	X	Neither	Both
	$N(\alpha_1)$	True	X		
$B = R$	$Y(\alpha_1)$	True	✔	Both	Neither
	$N(\alpha_1)$	False	✔		
$C = R$	$Y(\alpha_1)$	True	✔	Yes	No
	$N(\alpha_1)$	True	X		

As a consequence, True will answer λ with 'neither', while False will answer with 'both'. The answers to questions *sb* and τ are explained similarly. In section "Formalizations via Theories of Truth" we will return to this procedure in more detail. Let us now move forward to our solution to the puzzle.

Our two-question solution has the following structure. First, we ask a question which allows us to identify a god which is not Random. Then, we ask a follow up question to the god which we know not to be Random, and use the answer we get to determine the identity of all three gods.

Finding a god that is not Random

Our first question, α_1, is defined as follows:

$$\alpha_1 : (N(\alpha_1) \text{ and } A = R) \text{ or } (Y(\alpha_1) \text{ and } B = R) \text{ or } C = R$$

Table 2 investigates the consequences of answering α_1 with 'yes' or 'no' relative to the world under consideration (first column) and reports the reactions of True and False to α_1, which are a function of those consequences as we illustrated above.

Let's explain the first two rows. When A is Random and α_1 is answered with 'yes', α_1 is false—as all its three disjuncts are—and so answering α_1 with 'yes' is incorrect when A is Random. Similarly, when A is Random and α_1 is answered with 'no', α_1 is true and so when A is Random, answering α_1 with 'no' is incorrect as well. So, when A is Random, True will answer α_1 with 'neither', while False will answer it with 'both'. The other entries in the table are explained similarly. We address α_1 to A and extract the following information from his answers.

Conclusion 1 is only an intermediate stage for arriving at Conclusion 2, which, as a function of A's answer to α_1, states which god is not Random. Table 3 is, in combination with Table 2, self-explanatory.

Determining the identity of A, B and C by a follow up question

By asking question α_1 to A, we either learn that B is not Random or that C is not Random. We assume that we learn that B is not Random, the case where C is not Random being similar. As B is not Random, exactly one of the following four statements is true:

$$p_1 := B = T \text{ and } A = F \text{ and } C = R. \quad p_2 := B = T \text{ and } A = R \text{ and } C = F$$
$$p_3 := B = F \text{ and } A = T \text{ and } C = R. \quad p_4 := B = F \text{ and } A = R \text{ and } C = T$$

Table 3 Conclusions based on A's answer to α_1

A's answer	Conclusion 1	Conclusion 2
Yes	$(A = T$ and $C = R)$ or $A = R$	$B \neq R$
No	$(A = F$ and $C = R)$ or $A = R$	$B \neq R$
Neither	$(A = R$ and $A = T)$ or $(A = F$ and $B = R)$ or $A = R$	$C \neq R$
Both	$(A = R$ and $A = F)$ or $(A = T$ and $B = R)$ or $A = R$	$C \neq R$

Table 4 Reactions of True and False on α_2

World	Y/N	$\mathcal{V}(\alpha_2)$	Status Y/N	True	False
p_1	$Y(\alpha_2)$	False	X	Neither	Both
	$N(\alpha_2)$	True	X		
p_2	$Y(\alpha_2)$	True	✔	Both	Neither
	$N(\alpha_2)$	False	✔		
p_3	$Y(\alpha_2)$	True	✔	Yes	No
	$N(\alpha_2)$	True	X		
p_4	$Y(\alpha_2)$	False	X	No	Yes
	$N(\alpha_2)$	False	✔		

Table 5 Conclusions based on B's answer to α_2

B's answer	Conclusion 1	Conclusion 2
Yes	$(B = T$ and $p_3)$ or $(B = F$ and $p_4)$	p_4
No	$(B = T$ and $p_4)$ or $(B = F$ and $p_3)$	p_3
Neither	$(B = T$ and $p_1)$ or $(B = F$ and $p_2)$	p_1
Both	$(B = F$ and $p_2)$ or $(B = F$ and $p_1)$	p_2

We will ask B, whom we know not to be Random, question α_2:

$$\alpha_2 : (N(\alpha_2) \text{ and } p_1) \text{ or } (Y(\alpha_2) \text{ and } p_2) \text{ or } p_3$$

Table 4 has exactly the same rationale as Table 2.

Table 5, which has exactly the same rationale as Table 3, shows that B's answer to α_2 allows us to determine whether p_1, p_2, p_3 or p_4 is the case, which means that B's answer allows us to determine the identity of all three gods.

Gods Who Answer with 'da' and 'ja'

We will now solve $HLPE^4_{syn}$, in which the gods answer positively and negatively by using, in some order, the words 'da' and 'ja'. The methods of the previous section easily carry over to this slightly more complicated puzzle. Let $M(d, y)$ and $M(d, n)$ abbreviate "da' means 'yes" and "da' means 'no" respectively. Further, with x an arbitrary question, $D(x)$ reads as 'your answer to x is 'da", while $J(x)$ reads as 'your answer to x is 'ja".

Table 6 Reactions of True and False on β_1

World	Language	*D*/*J*	$\mathcal{V}(\beta_1)$	✔/X	True	False
$A = R$	$M(d, y)$	$D(\beta_1)$	True	✔	Both	Neither
		$J(\beta_1)$	False	✔		
$B = R$	$M(d, y)$	$D(\beta_1)$	False	X	Neither	Both
		$J(\beta_1)$	True	X		
$C = R$	$M(d, y)$	$D(\beta_1)$	True	✔	da	ja
		$J(\beta_1)$	True	X		
$A = R$	$M(d, n)$	$D(\beta_1)$	False	✔	Both	Neither
		$J(\beta_1)$	True	✔		
$B = R$	$M(d, n)$	$D(\beta_1)$	True	X	Neither	Both
		$J(\beta_1)$	False	X		
$C = R$	$M(d, n)$	$D(\beta_1)$	False	✔	da	ja
		$J(\beta_1)$	False	X		

Finding a god that is not Random

Our first question, β_1, is defined as follows:

$$\beta_1 : M(d, y) \text{ iff } ((D(\beta_1) \text{ and } A = R) \text{ or } (J(\beta_1) \text{ and } B = R) \text{ or } C = R)$$

In Table 6, we investigate the consequences of answering β_1 with 'da' or 'ja' relative to a world in which Random is *A*, *B* or *C* and to a language in which 'da' means either 'yes' or 'no'. The Table 6, reports the reactions of True and False to β_1, which are a function of the investigated consequences. Due to our uncertainty with respect to the meaning of 'da' and 'ja', Table 6 has 12 (rather than 6) rows. Let us compare row 1 with row 7. The first row tells us that when *A* is Random and 'da' means 'yes', answering β_1 with 'da' renders β_1 true. As on the first row 'da' means 'yes', answering 'da' to β_1 under the conditions of the first row is correct. Row 7 tells us that, when *A* is Random and 'da' means 'no', answering 'da' to β_1 renders β_1 false. Accordingly, answering 'da' to β_1 under the conditions of the seventh row is correct. From Table 6, it easily follows that asking β_1 to *A* allows us to determine the identity of a god which is not Random. Drawing the "conclusion table" associated with Table 6 is left to the reader.

Determining the identity of A, B and C by a follow up question

By asking question β_1 to *A*, we either learn that *B* is not Random or that *C* is not Random. Again, we assume that we learn that *B* is not Random, the case where *C* is not Random being similar. When *B* is not Random, exactly one of p_1, p_2, p_3 and p_4 is true. As a follow up question to β_1, we will ask β_2 to the non Random god *B*.

$$\beta_2 : M(d, y) \text{ iff } ((D(\beta_2) \text{ and } p_1) \text{ or } (J(\beta_2) \text{ and } p_2) \text{ or } p_3)$$

Table 7, describes the reactions of True and False to β_2 relative to the world and language under consideration. From Table 7, it follows that asking β_2 to *B*, which is not Random, allows us to determine the identity of all three gods. Drawing the "conclusion table" associated with Table 7 is left to the reader.

Table 7 Reactions of True and False to β_2

World	Language	*D*/*J*	$\mathcal{V}(\beta_2)$	✔/X	True	False
p_1	$M(d, y)$	$D(\beta_2)$	True	✔	Both	Neither
		$J(\beta_2)$	False	✔		
p_2	$M(d, y)$	$D(\beta_2)$	False	X	Neither	Both
		$J(\beta_2)$	True	X		
p_3	$M(d, y)$	$D(\beta_2)$	True	✔	da	ja
		$J(\beta_2)$	True	X		
p_4	$M(d, y)$	$D(\beta_2)$	False	X	ja	da
		$J(\beta_2)$	False	✔		
p_1	$M(d, n)$	$D(\beta_2)$	False	✔	Both	Neither
		$J(\beta_2)$	True	✔		
p_2	$M(d, n)$	$D(\beta_2)$	True	X	Neither	Both
		$J(\beta_2)$	False	X		
p_3	$M(d, n)$	$D(\beta_2)$	False	✔	da	ja
		$J(\beta_2)$	False	X		
p_4	$M(d, n)$	$D(\beta_2)$	True	X	ja	da
		$J(\beta_2)$	True	✔		

Formalizations via Theories of Truth

As noted in the introduction, all the previous solutions to *HLPE* are presented informally using natural language. In the previous section, we likewise introduced our four-valued conception of True and False informally by showing how it can be applied to solve $HLPE^3_{syn}$. In this section, we discuss the prospects of formally representing the present and previous solutions to *HLPE*. The behavior of the gods True and False in *HLPE* suggests that a formalization of their behavior can fruitfully be based upon a formal theory of truth. In this section, we follow this suggestion by basing ourselves upon Strong Kleene (Kripkean) fixed point theories of truth. To be sure, there are various theories of truth; we could also work with an account of True and False that is based on say, a revision theory of truth (cf. Gupta and Belnap 1993) or on fixed points that are constructed in accordance with the Supervaluation schema. We choose to work with Strong Kleene theories because such theories are very well-known, easy to present and, importantly, they allow us to represent the solutions to *HLPE* in a sense that will be made clear below.[5]

In fact, we will not apply our formal modeling to *HLPE* itself, but rather to *the four roads riddle*, presented in section "The Four Roads Riddle". The four roads riddle may be considered as a simplified version of *HLPE* while containing *HLPE*'s essential features: our formalization of the four roads riddle is easily seen to carry over to (versions of) *HLPE*. The formal language in which we will study the four roads riddle contains a 'yes' and a 'no' predicate, but no "non-standard" answer predicates, such as predicates for 'both', 'neither', 'silence' or what have you.

[5] Which is not to say that other theories of truth do not allow such representation.

As none of the solutions to *HLPE* involves questions that are formed using non-standard answer predicates, the expressive limitations of our language do not prevent us from representing these solutions. To be sure, ultimately one wants an account of the behavior of True and False in a more expressive language which does contain non-standard answer predicates. In section "Critical Remarks on Formalizations", we will briefly comment on the prospects of such an account.

After presenting the four roads riddle in sections "The Four Roads Riddle", and "Formalizations" is concerned with formalizations of the riddle. Section "Critical Remarks on Formalizations" critically looks back at what has been achieved in section "Formalizations". Section "The Wheeler and Barahona Argument" discusses the information theoretic argument of Wheeler and Barahona (2011) that was mentioned in the introduction.

The Four Roads Riddle

The Riddle

You arrive at a cross roads at which you can head either *north, south, east* or *west*. You know that only one of the four roads, call it the *good road*, leads to your destination. Unfortunately, you have no clue as to which road is good. However, two gods, call them a and b, are situated at the cross roads. You know that one of these gods is True while the other god is False, but you have no clue as to whether a or b is True. The four roads riddle is as follows. Given the circumstances just sketched, can you come up with a single question that, when posed to either one of the gods, allows you to determine which road is good?

The Language L_B and its Ground Models

We start out by introducing a restricted formal language in which we will study the four roads riddle. Our basic formal language is a quantifier free[6] predicate language with identity L_B, consisting of the following non-logical vocabulary.[7]

Constant symbols:

- a and b, which denote, in some order, **Tr**ue and **Fa**lse.
- g_T and g_F, which denote, respectively, **Tr**ue and **Fa**lse.
- n, w, e, s, which denote, respectively, the **no**rth, **we**st, **ea**st and **so**uth road.
- $\{[\sigma] \mid \sigma \in Sen(L_B)\}$: *quotational constant symbols*;[8] for each $\sigma \in Sen(L_B)$, $[\sigma]$ denotes σ.

[6] We do so for sake of simplicity: the definition of the three- and four-valued answering functions below are easily seen to carry over to quantified languages.

[7] We will use $=$, $\wedge$, $\vee$, $\neg$, $\rightarrow$ and $\leftrightarrow$ and as logical symbolism, the interpretation of which is as expected.

[8] The set of quotational constant symbols has a joint recursive definition together with $Sen(L_B)$, the set of sentences of L_B. The definition of these sets can safely be left to the reader.

- $C = \{c_1, c_2, \ldots, c_n\}$: *non-quotational constant symbols*, which denote (arbitrary) elements of $Sen(L_B)$ and which can be used to define self-referential sentences. [9]

Predicate symbols:

- $G(x)$, interpreted as 'x is the good road'.
- $Y(x, y)$ and $N(x, y)$, interpreted as 'the answer of x to y is 'yes'' and 'the answer of x to y is 'no'' respectively.

A *ground model* $M = (D, I)$ is an interpretation of the "yes/no predicate free fragment of L_B" which respects the intuitive interpretation of L_B that is given above. More precisely, a ground model $M = (D, I)$ is a classical model for $L_B^- = L_B - \{Y, N\}$ which respects the following clauses:

1. $D = \{\mathbf{Tr}, \mathbf{Fa}, \mathbf{no}, \mathbf{ea}, \mathbf{so}, \mathbf{we}\} \cup Sen(L_B)$
2. $I(g_T) = \mathbf{Tr}$, $I(g_F) = \mathbf{Fa}$, $I(n) = \mathbf{no}$, $I(e) = \mathbf{ea}$, $I(w) = \mathbf{we}$, $I(s) = \mathbf{so}$
3. $I([\sigma]) = \sigma$ for all $\sigma \in Sen(L_B), I(c_i) \in Sen(L_B)$ for all $c_i \in C$
4. Either ($I(a) = \mathbf{Tr}$ and $I(b) = \mathbf{Fa}$) or ($I(b) = \mathbf{Tr}$ and $I(a) = \mathbf{Fa}$)
5. Either $I(G) = \{\mathbf{no}\}$ or $I(G) = \{\mathbf{ea}\}$ or $I(G) = \{\mathbf{so}\}$ or $I(G) = \{\mathbf{we}\}$.

For any ground model M, we will use $\mathcal{C}_M : Sen(L_B^-) \rightarrow \{0, 1\}$ to denote the (classical) valuation of L_B^- that is induced by M. A ground model fixes all the relevant facts; facts about the world on the one hand and facts about sentential reference on the other. As such, an account of the behavior of True and False owes us an explanation of how True and False answer (arbitrary) L_B sentences relative to a ground model. Below, we are concerned with such explanations.

Formalizations

A Three-Valued Answering Function for L_B

Clearly, the predicates $Y(g_T, \cdot)$ and $N(g_T, \cdot)$ bear a close similarity with, respectively, a truth predicate and a falsity predicate. Similarly, the predicates $Y(g_F, \cdot)$ and $N(g_F, \cdot)$ bear a close similarity with, respectively, a falsity predicate and a truth predicate. When we treat our yes/no predicates as truth/falsity predicates in the sense alluded to, Kripke's fixed point techniques, as described in Kripke (1975), may be readily applied in the present setting. In this section, those techniques will be applied to define a three-valued answering function of True and False with an eye on satisfying the following two desiderata:

A (The construction of) the answering function allows us to represent the previous (three-valued) solutions to *HLPE*.

B The answering function gives the intuitive correct verdict with respect to L_B questions that are not considered in those solutions.

[9] For instance, when posed to god a, the sentence $Y(a, c_1)$ may be paraphrased as: 'Is it the case that: your answer to this question is 'yes'?', provided that the denotation of c_1 is $Y(a, c_1)$.

Here we go. By a *(Strong Kleene) fixed point valuation for L_B over a ground model M*, $\mathcal{K}_M : Sen(L_B) \rightarrow \{0, \frac{1}{2}, 1\}$, we mean a three-valued valuation of L_B which respects the following five clauses. Below, $\overline{\sigma}$ is an arbitrary constant of L_B (quotational or non-quotational) which denotes $\sigma \in Sen(L_B)$.

1. $\mathcal{K}_M(\sigma) = \mathcal{C}_M(\sigma)$ for all $\sigma \in Sen(L_B^-)$
   $\mathcal{K}_M$ respects the ground model M.
2. $\mathcal{K}_M(Y(g_T, \overline{\sigma})) = \mathcal{K}_M(\sigma), \mathcal{K}_M(N(g_T, \overline{\sigma})) = 1 - \mathcal{K}_M(\sigma)$ *Fixed point condition for $Y(g_T, \cdot)$ and $N(g_T, \cdot)$.*
3. $\mathcal{K}_M(Y(g_F, \overline{\sigma})) = 1 - \mathcal{K}_M(\sigma), \mathcal{K}_M(N(g_F, \overline{\sigma})) = \mathcal{K}_M(\sigma)$ *Fixed point condition for $Y(g_F, \cdot)$ and $N(g_F, \cdot)$.*
4. $\mathcal{K}_M(Y(t_1, t_2)) = \mathcal{K}_M(N(t_1, t_2)) = 0$, when $I(t_1) \notin \{\mathbf{Tr}, \mathbf{Fa}\}$ or $I(t_2) \notin Sen(L_B)$. *Only questions receive answers and only gods answer questions.*
5. (a) $\mathcal{K}_M(\neg\sigma) = 1 - \mathcal{K}_M(\sigma)$
   (b) $\mathcal{K}_M(\alpha \wedge \beta) = min\{\mathcal{K}_M(\alpha), \mathcal{K}_M(\beta)\}$
   (c) $\mathcal{K}_M(\alpha \vee \beta) = max\{\mathcal{K}_M(\alpha), \mathcal{K}_M(\beta)\}$ *$\mathcal{K}_M$ is Strong Kleene.*

In general, a ground model M allows us to define various fixed point valuations over it.[10] We could define an answering function for True and False that is based on, say, the minimal fixed point valuation over M or, say, the maximal intrinsic fixed point valuation. As will be clear from the discussion below, these answering functions allow us to represent previous solutions to *HLPE* (**A**) but, arguably, they do not give the intuitive correct verdict with respect to L_B questions that are not considered by those solutions (**B**). In order to justice to both **A** and **B**, we define the valuation function $\mathcal{K}_M^\star : Sen(L_B) \rightarrow \{0, \frac{1}{2}, 1\}$ by quantifying over all Strong Kleene fixed point valuations over M. $\mathcal{K}_M^\star$ is defined as follows, where the quantifiers range over all Strong Kleene fixed point valuations over M.

- $\mathcal{K}_M^\star(\sigma) = 1 \Leftrightarrow \exists \mathcal{K}_M : \mathcal{K}_M(\sigma) = 1 \;\&\; \nexists \mathcal{K}_M : \mathcal{K}_M(\sigma) = 0$
- $\mathcal{K}_M^\star(\sigma) = \frac{1}{2} \Leftrightarrow \nexists \mathcal{K}_M : \mathcal{K}_M(\sigma) = 1 \;\&\; \nexists \mathcal{K}_M : \mathcal{K}_M(\sigma) = 0$
- $\mathcal{K}_M^\star(\sigma) = 0 \Leftrightarrow \exists \mathcal{K}_M : \mathcal{K}_M(\sigma) = 0$

The valuation $\mathcal{K}_M^\star$ is used to define an answering function for True and False as follows.

Answering function based on $\mathcal{K}_M^\star$:

i. True (False) answers σ with 'yes' just in case $\mathcal{K}_M^\star(\sigma) = 1$ ($\mathcal{K}_M^\star(\sigma) = 0$).
ii. True (False) answers σ with 'no' just in case $\mathcal{K}_M^\star(\sigma) = 0$ ($\mathcal{K}_M^\star(\sigma) = 1$).
iii. True and False remain silent on σ just in case $\mathcal{K}_M^\star(\sigma) = \frac{1}{2}$.

Let us first point out why we choose to work with $\mathcal{K}_M^\star$ and not with, say, the minimal or maximal intrinsic fixed point. To do so, consider the following three questions:

[10] In the present setting, the number of fixed point valuations over M depends on the denotations of the members of C; if, say, $I(c) = (g_T = g_T)$ for every $c \in C$, there is a unique fixed point valuation over M.

Table 8 Values of $\mathcal{K}_M^\star$ for $I(\theta)$, $I(\tau)$, $I(\lambda)$

World	$\mathcal{K}_M^\star(I(\theta))$	$\mathcal{K}_M^\star(I(\tau))$	$\mathcal{K}_M^\star(I(\lambda))$
$a = g_T$	1	0	$\frac{1}{2}$
$a = g_F$	1	$\frac{1}{2}$	0

θ: Is your answer to θ 'yes' or 'no'?
λ: Is your answer to λ 'no'?
τ: Is your answer to τ 'yes'?

To remove the indexical 'your', we assume that the questions are addressed to god a. In order to represent the questions in L_B then, we let θ, λ and τ be non-quotational constants such that $I(\theta) = Y(a, \theta) \vee N(a, \theta)$, $I(\lambda) = N(a, \lambda)$ and $I(\tau) = Y(a, \tau)$. Table 8 describes how $\mathcal{K}_M^\star$ valuates these questions.

Consider question θ. First note that the $\mathcal{K}_M^\star$ account of True and False prescribes that True answer θ with 'yes' and that False answers θ with 'no'. I take it that this is how it, intuitively, should be.[11] This provides a reason for preferring the $\mathcal{K}_M^\star$ account of True and False above an account that is based on the minimal fixed point; as θ is *ungrounded*, the minimal fixed point will valuate it as $\frac{1}{2}$, implying that both True and False must remain silent on θ according to the minimal fixed point. To see how θ obtains its $\mathcal{K}_M^\star$ value, note that $Y(g_T, \cdot)$ and $N(g_F, \cdot)$ are truth predicates in disguise, whereas $Y(g_F, \cdot)$ and $N(g_T, \cdot)$ are disguised falsity predicates. Thus, when posed to True, question θ allows for the alethic paraphrase 'this very sentence is true or false', whereas, when addressed to False, the paraphrase becomes 'this very sentence is false or true'. Clearly then, there is a fixed point in which these sentences are true while there is no fixed point in which they are false; $\mathcal{K}_M^\star(I(\theta)) = 1$, irrespective of whether we address θ to True or False.

The maximal intrinsic fixed point also valuates θ as 1 and so an account of True and False based on it would prescribe the same answers to θ as the $\mathcal{K}_M^\star$ account. We prefer the $\mathcal{K}_M^\star$ account over the account based on the maximal intrinsic fixed point due to the answers that are prescribed to question τ. According to the $\mathcal{K}_M^\star$ account of True and False, True answers question τ with 'no', whereas False remains silent on τ. Intuitively—as also remarked in Rabern and Rabern (2008)—False must indeed remain silent on τ, as he cannot answer it "in accordance with his nature", which is to speak falsely. Although the previous solutions to *HLPE* do not discuss how True should answer τ, their authors do state that the gods remain silent on a question when they cannot answer that question "in accordance with their nature". But True clearly can answer τ with *either* 'yes' or 'no' "in accordance with his nature"—

[11] I take it that question θ reveals an interesting dissimilarity between positively answering a yes–no question and asserting its alethic counterpart: while 'yes' is clearly a *truthful* answer to θ, the ungroundedness of 'this very sentence is true or false' *may* deems its assertion inappropriate. More concretely, answering θ with 'yes' makes it true, while asserting 'this very sentence is true or false' does not render the asserted sentence true.

Table 9 Reactions of a to $Y(a, [G(n)])$

World	$\mathcal{K}_M^\star(Y(a,[G(n)]))$	Answer of a
$a = g_T$, $G(n)$	1	Yes
$a = g_T$, $\neg G(n)$	0	No
$a = g_F$, $G(n)$	0	Yes
$a = g_F$, $\neg G(n)$	1	No

although doing so is completely arbitrary—and so the question arises how True should answer τ. Now, one may take the arbitrariness of a yes/no answer to τ as a *further* reason for True to remain silent. However, this is not what the authors of previous solutions seem to have in mind.[12] So an account of True and False which prescribes that True answers τ with a yes/no answer seems more in line with the spirit of the previous solutions to *HLPE*. The $\mathcal{K}_M^\star$ account[13] is such an account, whereas an account based on the maximal intrinsic fixed point is not.

Note that, due to the relations between yes/no predicates and truth/falsity predicates, τ behaves like a Truthteller ('this very sentence is true') when addressed to True while it behaves like a Liar ('this very sentence is false') when addressed to False. As there is a fixed point in which the Truetheller is false, we get that $\mathcal{K}_M^\star(I(\tau)) = 0$ when a is True. As there is no fixed point in which the Liar is true and no fixed point in which the Liar is false, we get that $\mathcal{K}_M^\star(I(\tau)) = \frac{1}{2}$ when a is False. The $\mathcal{K}_M^\star$ valuation of question λ receives a dual explanation.

Putting $\mathcal{K}_M^\star$ to Work

Suppose that—in the setting of the four roads riddle—we (only) want to find out whether or not the north road is good. Asking the question 'is the north road good?' is useless; we do not know whether we address True or False when asking a question. However, a little reflection shows that the following question, when addressed to, say, god a, allows us to find out whether or not the north road is good:

Is your answer to the question 'is the north road good?' 'yes'? (1)

The L_B translation of question (1) is given by the sentence $Y(a, [G(n)])$. Table 9 explains, in terms of $\mathcal{K}_M^\star$, why asking question (1) suffices to find out whether or not the north road is good.

The table explains that 'yes' indicates that the north road is good and that 'no' indicates that the north road is not good. Question (1) is an instance of what is called the Embedded Question Lemma (**EQL**) in Rabern and Rabern (2008).

[12] Rabern and Rabern (2009) comment on the answering function that they had in mind in their published paper: according to this function, True gives a classical (yes/no) answer to questions like τ.

[13] Although the $\mathcal{K}_M^\star$ account prescribes that True answers τ with 'no', we do not think that there is any further reason to prefer such an account over an account according to which True answers τ with 'yes'. Further, some obvious modifications to $\mathcal{K}_M^\star$ will yield just such an account.

EQL Let E be the function that takes a question Q to the question 'Is your answer to the question 'Q' 'yes'?' When either True or False are asked $E(Q)$, an answer of 'yes' indicates that Q whereas an answer of 'no' indicates that not Q.

Proof Both a double positive and a double negative make a positive. □

Suppose that you addressed question (1) to a and that you received 'no' as an answer. So, now you know that either the south, east or west road is good—whereas you don't know whether a is True or False. Hence, we are left with the "three roads riddle". Next, we will show how to solve the three roads riddle via a single question, ρ, that is similar to the (crucial) questions that are exploited by previous (informal) self-referential solutions to *HLPE*. In the spirit of those solutions, we define ρ by referring to ρ in the argument place of the embedding function E of the **EQL**:

ρ : E((Is your answer 'no' to ρ and the south road is good) or the west road is good)

Questions like ρ, which refer to themselves in the argument place of the embedding function E, we call *self-embedded questions*. Note that the solution to $HLPE^3_{syn}$ that was given in the previous section does *not* rely on self-embedded questions. We'll return to this observation in section "'Both', 'Neither' and Self-Embedded Questions". In order to explain why the answer to ρ allows us to find out which of the three roads is good, it is convenient to first translate it into L_B. To do so, we let ρ be a non-quotational constant whose denotation is as follows:

$$Y(a, [(N(a, \rho) \wedge G(s)) \vee G(w)])$$

Here is an intuitive explanation of why ρ does the job. If west is the good road, the embedded question, i.e., $(N(a, \rho) \wedge G(s)) \vee G(w)$, will be true. Hence, True will answer the embedded question with 'yes' and False will answer it with 'no'. Thus, when asked whether they answer the embedded question with 'yes', i.e., when asked ρ, True and False will both answer with 'yes'. Similar reasoning shows that if east is the good road, True and False will both answer with 'no'. Finally, when south is the good road, question ρ reduces to a question which has the same answerhood conditions as the self-embedded question ρ_1:

ρ_1 : Is your answer 'yes' to the question of whether your answer to ρ_1 is 'no'?

As observed by Uzquiano (2010), neither True nor False can answer ρ_1 in accordance with his nature; they must remain silent on ρ_1. Similarly, when south is the good road, both True and False must remain silent on ρ. Table 10—whose construction can safely be left to the reader—shows that our $\mathcal{K}^{\star}_{M}$ based answering function yields the same verdict with respect to the answers of True and False to $I(\rho) = Y(a, [(N(a, \rho) \wedge G(s)) \vee G(w)])$.

Table 10 reveals the sense in which the $\mathcal{K}^{\star}_{M}$ account of True and False allows us to give a formal representation of the informal solution to the "three roads riddle". The principles at work in the solution to the "three roads riddle" are similar to the principles at work in the previous self-referential solutions to *HLPE*. Accordingly, the $\mathcal{K}^{\star}_{M}$ account of True and False can be used to represent these solutions as well.

Table 10 Reactions of a to $I(\rho)$

World	$\mathcal{K}^{\star}_{M}(I(\rho))$	Answer of a
$a = g_T$, $G(w)$	1	Yes
$a = g_T$, $G(e)$	0	No
$a = g_T$, $G(s)$	$\frac{1}{2}$	Silence
$a = g_F$, $G(w)$	0	Yes
$a = g_F$, $G(e)$	1	No
$a = g_F$, $G(s)$	$\frac{1}{2}$	Silence

Table 11 Values of $\mathcal{K}^{\bullet}_{M}$ for $I(\theta)$, $I(\tau)$, $I(\lambda)$

World	$\mathcal{K}^{\bullet}_{M}(I(\theta))$	$\mathcal{K}^{\bullet}_{M}(I(\tau))$	$\mathcal{K}^{\bullet}_{M}(I(\lambda))$
$a = g_T$	1	+	−
$a = g_F$	1	−	+

Putting a Four-Valued Answering Function for L_B to Work

We start by defining a four-valued valuation function of $L_B, \mathcal{K}^{\bullet}_{M} : Sen(L_B) \rightarrow \{0, +, -, 1\}$, in a similar manner as we defined $\mathcal{K}^{\star}_{M}$, i.e., by quantifying over all Strong Kleene fixed point valuations. We then define a four-valued answering function for True and False based on $\mathcal{K}^{\bullet}_{M}$ and show how it can be invoked to give a formal representation of a solution to the four roads riddle. The principles at work in our solution to the four roads riddle are similar to the principles at work in our solution to $HLPE^{4}_{syn}$ that was presented in section "Solving the Puzzles". Hence, the $\mathcal{K}^{\bullet}_{M}$ account can also be used to give a formal representation of our solution to $HLPE^{4}_{syn}$.

Here is the definition of $\mathcal{K}^{\bullet}_{M}$:

- $\mathcal{K}^{\bullet}_{M}(\sigma) = 1 \Leftrightarrow \exists \mathcal{K}_M : \mathcal{K}_M(\sigma) = 1 \;\&\; \nexists \mathcal{K}_M : \mathcal{K}_M(\sigma) = 0$
- $\mathcal{K}^{\bullet}_{M}(\sigma) = - \Leftrightarrow \nexists \mathcal{K}_M : \mathcal{K}_M(\sigma) = 1 \;\&\; \nexists \mathcal{K}_M : \mathcal{K}_M(\sigma) = 0$
- $\mathcal{K}^{\bullet}_{M}(\sigma) = + \Leftrightarrow \exists \mathcal{K}_M : \mathcal{K}_M(\sigma) = 1 \;\&\; \exists \mathcal{K}_M : \mathcal{K}_M(\sigma) = 0$
- $\mathcal{K}^{\bullet}_{M}(\sigma) = 0 \Leftrightarrow \nexists \mathcal{K}_M : \mathcal{K}_M(\sigma) = 1 \;\&\; \exists \mathcal{K}_M : \mathcal{K}_M(\sigma) = 0$

Table 11 illustrates how $\mathcal{K}^{\bullet}_{M}$ valuates questions θ, τ and λ, i.e., here is the $\mathcal{K}^{\bullet}_{M}$ version of Table 7.

According to the four-valued conception of True and False put forward in this paper, True answers τ with 'both', whereas False answers τ with 'neither'. Similarly, according to this conception, True answers λ with 'neither', whereas False answers λ with 'both'. This suggests the following answering function:

Answering function based on $\mathcal{K}^{\bullet}_{M}$:

i. True (False) answers σ with 'yes' just in case $\mathcal{K}^{\bullet}_{M}(\sigma) = 1 (\mathcal{K}^{\bullet}_{M}(\sigma) = 0)$.
ii. True (False) answers σ with 'no' just in case $\mathcal{K}^{\bullet}_{M}(\sigma) = 0 (\mathcal{K}^{\bullet}_{M}(\sigma) = 1)$.
iii. True and False answer σ with 'neither' just in case $\mathcal{K}^{\bullet}_{M}(\sigma) = -$.

iv. True and False answer σ with 'both' just in case $\mathcal{K}^{\bullet}_{M}(\sigma) = +$.

Suppose that that we want to solve the four roads riddle. It is not hard to calculate that, using the methods of section "Solving the Puzzles", question γ does the job, where γ denotes:

$$a = g_T \leftrightarrow (N(a, \gamma) \wedge G(n)) \vee (Y(a, \gamma) \wedge G(s)) \vee G(w)$$

By applying the methods of section "Solving the Puzzles"—which can safely be left to the reader—we see that an answer of 'yes' indicates that west is the good road, 'no' indicates that east is good, 'neither' indicates that north is good and 'both' indicates that south is good. The $\mathcal{K}^{\bullet}_{M}$ based answering function True and False yields exactly the same verdicts. To see why, we will consider (only) the case where the north road is good. So, suppose that the north road is good and that you address γ to True. Under these circumstances, the answerhood conditions of γ are equivalent to the answerhood conditions of the following question:

$$\gamma_1 : g_T = g_T \leftrightarrow N(g_T, \gamma_1)$$

As the left-hand side of γ_1 is true, γ_1 is true just in case its right-hand side, i.e., $N(g_T, \gamma_1)$ is true. But $N(g_T, \cdot)$ functions as a falsity predicate and so, $N(g_T, \gamma_1)$ is true just in case γ_1 is false; we get that γ_1 is true just in case it is false. Hence, γ_1 is *paradoxical*. In other words, when a is True and the north road is good, we get that $\mathcal{K}^{\bullet}_{M}(I(\gamma_1) = -$ and so True replies γ_1 with 'neither'. Now suppose that the north road is good and that you address γ to False. Under these circumstances, the answerhood conditions of γ are equivalent to the answerhood conditions of the following question:

$$\gamma_2 : g_F = g_T \leftrightarrow N(g_F, \gamma_1)$$

As the left-hand side of γ_2 is false, γ_2 is true just in case its right-hand side, i.e., $N(g_F, \gamma_2)$ is false. But $N(g_F, \cdot)$ functions as a truth predicate and so, $N(g_F, \gamma_2)$ is false just in case γ_2 is false; we get that γ_2 is true just in case it is false. Hence, γ_2 is *paradoxical*. In other words, when a is False and the north road is good, we get that $\mathcal{K}^{\bullet}_{M}(I(\gamma_1) = -$ and so False replies γ_2 with 'neither'. So, when the north road is good, both True and False reply to γ with 'neither'. The other three cases are reasoned out similarly and so the answers of True (and False) to γ as obtained according to the method of section "Solving the Puzzles" are the same as the answers that are obtained via the $\mathcal{K}^{\bullet}_{M}$ based answering function. Similarly, one can show that the solution to $HLPE^{4}_{syn}$ that was presented in section "Solving the Puzzles", allows for a formal representation using our $\mathcal{K}^{\bullet}_{M}$ based answering function.

Note that our solution to the four roads riddle, i.e., question γ, is not a *self-embedded question*. Likewise, none of the questions discussed in section "Solving the Puzzles" are self-embedded. In contrast, our solution to the three roads riddle, i.e., question ρ, is a self-embedded question and so are the (crucial) questions invoked in previous self-referential solutions to *HLPE*. In the next section, we will explain, amongst others, in which sense self-embedded questions give rise to a problem for the intuitive interpretation of the answers 'both' and 'neither' that was sketched in the introduction.

Critical Remarks on Formalizations

'Both', 'Neither' and Self-Embedded Questions

In section "Putting $\mathcal{K}_M^\star$ to Work", we discussed the self-embedded question ρ_1, whose L_B translation is as follows:

$$\rho_1 : Y(a, [N(a, \rho_1)])$$

Due to the self-embedding that is present in ρ_1, it is not clear how the methods of section "Solving the Puzzles" should be applied to calculate the answers of True and False to it; a yes/no answer to ρ_1 does not (immediately) render the statement true or false. However, it is clear how $\mathcal{K}_M^\bullet$ valuates ρ_1. Exploiting the similarity between yes/no predicates and truth/falsity predicates, we see that, when addressed to True, ρ_1 may be paraphrased in alethic terms as 'it is true that this very sentence is false'. When addressed to False, the paraphrase becomes 'it is false that this very sentence is true'. Clearly then, we have that $\mathcal{K}_M^\bullet(I(\rho_1)) = -$, irrespective of whether we address ρ_1 to True or False. But this means that both True and False will answer ρ_1 by replying 'ρ_1 can *neither* be answered with 'yes' or 'no''. Observe that this $\mathcal{K}_M^\bullet$ prescription is at odds with our original interpretation of the answers 'neither' and 'both', according to which True, in answering 'neither' to question λ speaks truly, whereas False, in answering λ with 'both' speaks falsely. Indeed, as ρ_1 can neither (on pain of a self-contradiction) be answered with 'yes' or 'no', False, in replying with 'neither' cannot be said to answer ρ falsely.

So, although the method of section "Solving the Puzzles" does not prescribe how the answers to ρ_1 should be calculated, the answers to ρ_1 that are prescribed by $\mathcal{K}_M^\bullet$ do not fit in with intended interpretation of 'both' and 'neither'. Thus, two options suggest themselves, which are associated with two distinct conceptions of True and False:

1. Stick to the intended interpretation of 'both' and 'neither' and extend the method of section "Solving the Puzzles" such that it becomes applicable to self-embedded questions and such that, in particular, the answer given by False to ρ_1 is 'both'. This option is naturally associated with an *informative conception* of True and False in which, in answering our questions, they intend to convey information. For instance, in answering a Liar question λ with 'neither' True intends to convey the information that he can't answer λ with 'yes' or 'no'.
2. Use the $\mathcal{K}_M^\bullet$ account of True and False and give up the interpretation of 'both' and 'neither'. For instance, say that if $\mathcal{K}_M^\bullet(\sigma) = -$, True and False reply to σ with an explosion, while if $\mathcal{K}_M^\bullet(\sigma) = +$, they remain silent. On this account, the non-linguistic actions of exploding and remaining silent have their origin in the "paradoxality" and the "arbitrariness" of the possible yes/ no answers. Being non-linguistic actions, exploding and remaining silent are not to be evaluated in terms of 'speaking truly' and 'speaking falsely'. This option is naturally associated with an *algorithmic conception* of True and False in which, in answering our questions, they do not intend to convey any information, but

rather, they follow an algorithm. Explosions and silences, on this conception, are best thought of as two distinct ways in which the algorithm can fail, due to the non-existence of solutions (paradoxality) and the abundance of solutions (arbitrariness) respectively. On the algorithmic conception of True and False, explosions and silences are not genuine *answers*, but rather, states that the gods end up in due to their processing of certain questions.

In a sense, it is more natural to speak of the failure of an algorithm due to the lack of solutions (paradoxality) than due to the lack of the abundance of solutions (arbitrariness). As such, the algorithmic conception of True and False is, arguably, more naturally associated with the 3-valued account of True and False that is underlying the previous solutions to *HLPE*.[14]

I take it that option 1 is preferable; I take it that an account of True and False according to which these gods can be understood as always speaking, respectively, truly and falsely, is preferable over an account on which they sometimes do not speak at all. Such an account simply seems to do more justice to Boolos' remarks that 'True *always* speaks truly' and 'False *always* speaks falsely'. To be sure, Boolos may not have envisioned the possibility to ask self-referential questions. Then again, I take it that an account of True and False which manages to respect Boolos' instructions, even in the presence of self-reference, is preferable to an account which does not.

Although it is beyond the scope of this paper to carry out option 1 in a rigorous way, here is a hint of how one may proceed. The crucial aspect of the method of section "Solving the Puzzles" was that True and False calculate how their yes/no answers to a question σ influence the truth-value of σ, in light of which they judge these yes/no answers to be correct/incorrect. Based on those judgements, they then decide which answer they actually give to σ. So according to the method of section "Solving the Puzzles", the patterns of reasoning of True and False leading up to the correct/incorrect judgement are exactly the same; they only differ in how these judgements are converted into answers. In particular, with respect to questions like λ and τ, True and False find exactly the same judgements. In a sense, this means that False first calculates whether answering with yes/no is *objectively* correct/incorrect and then decides to lie about these judgements. This idea, of False first calculating whether a yes/no answer is *objectively* correct and *then* lying about his findings, can be put to work in extending the method of section "Solving the Puzzles" to bear on self-embedded questions. Table 12 which will be used to explain how True (and False) calculate their answer to ρ_1 in this manner.

Let us explain. The first row supposes that ρ_1 is answered with 'yes' (by a). As a consequence, the embedded question '$N(a, \rho_1)$' is false, as indicated in the second column. But this means that the *objectively correct* answer to the embedded question is 'no'. Accordingly, $Y(a, [N(a, \rho_1)])$—which may here be thought of as

[14] As pointed out by an anonymous referee, it can be argued that a 3-valued algorithmic conception of True and False does not require that we introduce a predicate in our language that represents failures of the algorithm as such failures do not belong to the language of the gods. In other words, it can be argued that the problem of *expressive completeness* (see section "Expressive Incompleteness") does not arise on a 3-valued algorithmic conception of True and False.

Table 12 Reactions of True and False on ρ_1

Y/N	$\mathcal{V}(N(a, \rho_1))$	$\mathcal{V}(Y(a, [N(a, \rho_1)]))$	✔/X	True	False
$Y(a, \rho_1)$	False	False	X	Neither	Both
$N(a, \rho_1)$	True	True	X		

'the answer that *should* be given to '$N(a, \rho_1)$' is 'yes"—is false, as indicated in the third column. Hence, as $I(\rho_1) = Y(a, [N(a, \rho_1)])$, answering 'yes' to ρ_1 is incorrect. The second row receives a similar explanation. Accordingly, True answers ρ_1 with 'neither' while False answers with 'both'. So according to the envisioned method for processing self-embedded questions, False finds out whether answering with yes/no is objectively correct—and not, say "correct for False"—and lies about his findings. Using exactly the same principles, the answer to "deeper" embedded questions, such as ρ_2, can be calculated.

$$\rho_2 : Y(a, [N(a, [Y(a, \rho_2)])])$$

Although these remarks on self-embedded questions do not define a rigorous algorithm for calculating the answers of True and False, I take it that they illustrate that, despite our possibility to ask self-embedded questions, there are hopes for developing a formal account of True and False according to which they can be understood in accordance with Boolos' instructions. However, self-embedded questions aside, a satisfying formal account of True and False faces more issues that have to be resolved. Below we discusses two such issues.

Expressive Incompleteness

As noted before, none of the solutions to *HLPE* exploits *non-standard questions*, i.e., questions that are formed with "non-standard answer predicates". In particular, L_B only contains answer predicates associated with 'yes' and 'no' and, in that sense, L_B may be called *expressive incomplete*. Although none of the solutions exploits non-standard questions, it seems reasonable to ask how True and False answer such questions. For instance, how do True and False answer questions like:

μ_1 : Is your answer to μ_1 'no' or 'neither'?

μ_2 : Do you answer 'yes' to the question of whether you answer 'neither' to μ_2?

Theories of truth typically do not contain predicates associated with the non-classical semantic values they employ in their meta-language. As such, we cannot expect much guidance from theories of truth in developing a formal account of how True and False answer non-standard questions. However, the methods of section "Solving the Puzzles" do give us some guidance here. Using '$NE(x, y)$' to abbreviate x answers y with 'neither', we can translate questions μ_1 and μ_2 by letting $I(\mu_1) = N(a, \mu_1) \vee NE(\mu_1)$ and $I(\mu_2) = Y(a, [NE(a, \mu_2)])$. Tables 13 and 14 explain how True and False answer μ_1 and μ_2.

Table 13 Reactions of True and False on μ_1

Y/N	$\mathcal{V}(N(a,\mu_1) \vee NE(\mu_1))$	✔/X	True	False
$Y(a, \mu_1)$	False	X	Neither	Both
$N(a, \mu_1)$	True	X		

Table 14 Reactions of True and False on μ_2

Y/N	$\mathcal{V}(NE(a,\mu_2))$	$\mathcal{V}(Y(a,[NE(a,\mu_2)]))$	✔/X	True	False
$Y(a, \mu_2)$	False	False	X	No	Yes
$N(a, \mu_2)$	False	False	✔		

Table 13 is self-explanatory. Note that, as True answers μ_1 with 'neither', μ_1 is true. Still, True does not answer μ_1 with 'yes' as in doing so he can be accused of lying. This situation with respect to μ_1—despite being true not being answered with 'yes' by True—has a clear rationale in terms of truthfully answering yes–no questions. However, it also points to a further[15] dissimilarity between yes–no questions and their alethic counterparts. For, consider μ_1', which is the alethic counterpart of μ_1:

$$\mu_1' : \text{Sentence } \mu_1' \text{ is false or (neither true nor false)}$$

In treating μ_1 and μ_1' alike, we are bound to conclude that μ_1' is 'neither true nor false'. But this exactly *what* μ_1' *says*, and so μ_1' is true and so *not* 'neither true nor false'. In sum, accepting that μ_1' is 'neither true nor false' seems to be tantamount to accepting a contradiction, while accepting that True answers μ_1 with 'neither' has a clear rationale in the (assumed) nature of True.

Table 14 explains how question μ_2, which is a self-embedded question, is answered. Table 14 is to be understood along the lines of section "Tokens or Types?". On the first line of Table 14, the consequences of answering with 'yes' are considered. Answering μ_1 with 'yes' renders $NE(a, \mu_2)$ false, which ensures that the correct answer to $NE(a, \mu_2)$ is 'no'. Accordingly, $Y(a, [NE(a, \mu_2)]$ is false and so answering μ_2 with 'yes' is incorrect.

Tokens or Types?

Consider the following two questions and suppose that we both address them to True.

$$\lambda : \text{Is your answer to } \lambda \text{ 'no'?}$$
$$\lambda_1 : \text{Is your answer to } \lambda \text{ 'no'?}$$

[15] For further dissimilarities, see footnote 11 and section " Tokens or Types?".

Question λ is familiar: True answers it with 'neither'. Question λ_1 asks whether True answers question λ with 'no'. As True answers λ with 'neither' (hence, *not* with 'no'), the truthful answer to λ_1 is 'no'. Hence, True should answer λ_1 with 'no'. Or so it seems. Yet if we base our account of True and False on, say $\mathcal{K}^{\bullet}_{M}$, we get different predictions. Fixed point theories of truth satisfy what is called the *intersubstitutability of truth*.[16] As a consequence, $Y(g_T, \overline{\sigma})$ and σ have the same semantic value according to $\mathcal{K}^{\bullet}_{M}$, whenever $\overline{\sigma}$ denotes σ. In particular then, we have that $\mathcal{K}^{\bullet}_{M}(Y(g_T, \lambda)) = \mathcal{K}^{\bullet}_{M}(Y(g_T, \lambda_1)) = -$ and so according to the $\mathcal{K}^{\bullet}_{M}$ account, True will answer both λ and λ_1 with 'neither'.

It is often argued that theories of truth *should* satisfy the intersubstitutability of truth, cf. Field (2008) or Beall (2009). In a nutshell, the argument is that if truth does not satisfy the intersubstitutability property, it can not play its stereotypical role of serving as a device of generalization. Let's accept this argument pertaining to theories that describe the behavior of a truth (and falsity) predicate. Does it carry over to a theory that describes the behavior of a 'answers with *yes*' (and 'answers with *no*') predicate? Not necessarily. For one thing, it is not clear that the stereotypical role of a 'answers with *yes*' predicate is to serve as a device of generalization and so the typical argument for the intersubstitutablity breaks down: although the analogy between true/falsity predicates and yes/no predicates is close, it is not perfect, as we also noted in the previous subsection. I take the intuitive reasoning with respect to λ and λ_1 that was given above convincing and I do not see why the intersubstitutablity of *truth* should lead us to dismiss that reasoning. Accordingly, I take it that a fully satisfying formal account of the behavior of True and False should be token-sensitive.

As the reader may have noticed, questions λ and λ_1 constitute an interrogative version of what Gaifman (1992) calls the "two lines puzzle". In fact, Gaifman's *pointer semantics* is an example of a token-sensitive theory of truth which gives up the intersubstitutability property and which yields (in alethic terms) similar conclusions with respect to the status of λ and λ_1 as the intuitive reasoning above. As such, it seems promising to develop a token-sensitive account of True and False on the basis of Gaifman's work. Clearly, doing so is far beyond the scope of this paper.

The Wheeler and Barahona Argument

Wheeler and Barahona (2011) argued that $HLPE^{3}_{syn}$ cannot be solved in less than three questions. Their argument relies on the following lemma from Information Theory.

> (*QL*) If a question has n possible answers, these answers cannot distinguish $m > n$ different possibilities.

Using *QL*, we see that when True and False have a three-valued answering repertoire (as in section "Critical Remarks on Formalizations"), we cannot solve the

[16] Meaning that $T(\overline{\sigma})$ and σ are intersubstitutable (without change of semantic value) in every non opaque context.

four roads riddle by asking a single question. In a similar vein—though in a more complicated setting—Wheeler and Barahona appealed to *QL* to argue that $HLPE^{3}_{syn}$ cannot be solved in less than 3 questions, where they assumed that True and False have a three-valued answering repertoire.

When True and False have a four-valued answering repertoire however, *QL* tells us that we *may* be able to solve the four roads riddle by asking a single question. Of course, whether or not we are actually able to do so depends on our ability to find questions such that their four possible answers are correlated to the four relevant states of affairs in a suitable way. Section "The Wheeler and Barahona Argument" showed how to solve the four roads riddle by a single question. So, by moving from a three- to a four-valued answering repertoire, we can escape the *QL* based conclusion that "the four roads riddle cannot be solved in a single question". Similarly, by moving from a three- to a four-valued answering repertoire, we escaped the conclusion of Wheeler and Barahona (2011) that $HLPE^{3}_{syn}$ cannot be solved in less than three questions.

Clearly then, the number of questions that is needed to solve "Smullyan like riddles" crucially depends on the number of answers that True and False have available. For instance, *QL* establishes that the *five roads riddle*—which is defined just as you expect it to be—cannot be solved (in one question) when True and False have a four-valued answering repertoire. However, *QL* leaves open the possibility that the five roads riddle can be solved in a setting in which True and False are assumed to have a five-valued answering repertoire. Here is such a setting. Assume that True and False are not omniscient and that, besides answering with 'yes', 'no', 'both' and 'neither', they *remain silent* when they are asked a question to which they do not know the answer. So, they now have a five-valued answering repertoire. Here is how to solve the five roads riddle. Let $p_1, \ldots, p_5$ be five sentences such that p_i states that the ith road is good and let $p_?$ be an unknowable (by True and False) sentence. The answer to question π, whose structure—a biconditional flanked by an atomic statement and a statement in disjunctive normal form—mirrors the structure of question γ of section "Putting a Four-Valued Answering Function for L_B to Work", allows you to find out which of the five roads is good:

$$\pi : a = g_T \leftrightarrow (N(a,\pi) \wedge p_1) \vee (Y(a,\pi) \wedge p_2) \vee (p_? \wedge p_3) \vee p_4$$

When p_3 is false, True and False know that $(p_? \wedge p_3)$ is false and so, depending on whether p_1, p_2, p_4 or p_5 is true, question π receives a similar treatment as question γ: the answers 'neither', 'both', 'yes' and 'no' are received just in case, respectively, p_1, p_2, p_4 and p_5 is true. However, when p_3 is true, the answerhood conditions of π reduce to those of $a = g_T \leftrightarrow (p_? \wedge p_3)$. As True and False do not know the truth value of $p_?$, they do not know the truth value of $(p_? \wedge p_3)$ and so they do not know the truth value of $a = g_T \leftrightarrow (p_? \wedge p_3)$. As a consequence, they must remain silent on π. Puzzle solved.

Let me conclude this section by stating a puzzle, $HLPE^{4*}_{syn}$, which I do not know how to solve (in two questions) and which is not unsolvable on the basis of *QL*. $HLPE^{4*}_{syn}$ is defined just like $HLPE^{4}_{syn}$, apart from the following difference. The gods react to your questions with 'huh', 'duh', 'da' and 'ja'. As before, 'da' and 'ja'

mean, in some order, 'yes' and 'no'. But now 'huh' and 'duh' mean, in some order, 'neither' and 'both'. Can we solve $HLPE^{4*}_{syn}$ in two questions?

Concluding Remarks

We put forward an alternative conception of how True and False answer yes–no questions, resulting in a four-valued answering repertoire. We then showed how this conception could be invoked to solve $HLPE^{4}_{syn}$ (and $HLPE^{3}_{syn}$) in two questions. Our four-valued (in contrast to a three-valued) answering repertoire allowed us to escape the argument of Wheeler and Barahona (2011), which established that $HLPE^{3}_{syn}$ cannot be solved in less than three questions.

The second part of the paper was concerned with formalizations of (the present and previous) solutions to versions of *HLPE*, that were all presented informally. We showed how—by appealing to Strong Kleene fixed point theories of truth and by working in a restricted setting—to give a formal representation of the solutions to *HLPE*. Although in an important sense our formalization "gets the job done", we discussed some desiderata of a formalization of the behavior of True and False that were not met by the one that was presented. To develop a formal account of True and False that meets these desiderata—i.e., a token-sensitive account for an expressive complete language in which True and False can be understood as, respectively, "always speaking truly" and "always speaking falsely"—is postponed to future work.

Acknowledgments Many thanks to Reinhard Muskens, Harrie de Swart and two anonymous referees of this journal for their valuable comments on this paper.

References

Beall, J. (2009). *Spandrels of truth*. Oxford: Oxford University Press.
Boolos, G. (1996). The hardest logic puzzle ever. *The Harvard Review of Philosophy, 6*, 62–65.
Field, H. (2008). *Saving truth from paradox*. Oxford: Oxford University Press.
Gaifman, H. (1992). Pointers to truth. *Journal of Philosophy, 89*(5), 223–261.
Gupta, A., & Belnap, N. (1993). *The revision theory of truth*. Cambridge: MIT Press.
Kripke, S. (1975). Outline of a theory of truth. *Journal of Philosophy, 72*, 690–716.
Rabern, B., & Rabern, L. (2008). A simple solution to the hardest logic puzzle ever. *Analysis, 68*, 105–112.
Rabern, B., & Rabern, L. (2009). In defense of the two question solution to the hardest logic puzzle ever. Unpublished.
Roberts, T. (2001). Some thoughts about the hardest logic puzzle ever. *Journal of Philosophical Logic, 30*, 609–612.
Uzquiano, G. (2010). How to solve the hardest logic puzzle ever in two questions. *Analysis, 70*, 39–44.
Wheeler, G., & Barahona, P. (2011). Why the hardest logic puzzle ever cannot be solved in less than three questions. *Journal of Philosophical Logic* (to appear).