A general form on the logic puzzles of Boolos*

Ken-etsu Fujita (Gunma University)
Toshihiko Kurata (Hosei University)
February 17, 2022

Abstract

Boolos (1996) posed the puzzle "The hardest logic puzzle ever" which had been devised by Raymond Smullyan, and gave a solution in the style of biconditional questions. We introduce a simple formalization of the puzzle consisting of questions, answerers, and answers in terms of propositional logic, and show its adequacy by the truth values $(0, 1)$ semantics. Then it turns out that the hardest logic puzzle can be considered as a natural extension of the puzzles of knights and knaves, i.e., lying and truth-telling by Smullyan. Here, we pose a general form of the puzzle, and provide partial solutions to some of the instances.

1 Introduction

George Boolos (1996) posed the puzzle "The hardest logic puzzle ever" which had been devised by Raymond Smullyan. To begin with, we quote the puzzle from Boolos [1]:

"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 Random speaks truly or falsely is a completely *random* matter. Your task is to determine the identities of A, B, 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.* $\cdots$ $\cdots$ $\cdots$"

He gave a solution in the style of biconditional questions as well.

1. Ask god A:

   does da means yes iff, you are True iff B is Random?

*This work was supported by the Research Institute for Mathematical Science, a Joint Usage/Research Center located in Kyoto University. This work was partly supported by Grants-in-Aid for Scientific Research KAKENHI (C) 17K05343 and 20K03711.

2. Ask B or C[1]:

   does da mean yes iff Rome is in Italy?

3. Does da mean yes iff A is Random?

Later, Roberts [5] and Rabern-Rabern [4] provided a simpler solution in the style of embedded questions as follows.

1. Ask god A:

   if I asked you if god B was Random, would you say da?

2. Ask B or C[1]:

   if I asked you if you always told the truth, would you say da?

3. If I asked you if god A was Random, would you say da?

We briefly review our methodology [2] in the next section. The methodology for solving the puzzles of Knights and Knaves from the book Logical Labyrinths [3] can be naturally extended to that for the hardest logic puzzle ever.

This elementary puzzle is known as a good example for the study of logic or sociology of *lying* and *truth-telling*. There were the island of knights and knaves such that *knights* always tell the truth, and *knaves* always lie. Each inhabitant of the island is either a knight or a knave.

- Problem 1.3. [3]:

  The island has two inhabitants, A and B. Now, A made the following statement:

  "Both of us are knaves."

  What is A and what is B?

From the book, we review the simple solution to this puzzle. Let A, B be propositional variables which mean that A is a knight and B is a Knight, respectively. Then $\neg A$ means that A is *not* a knight (i.e., knave) from the definition of knights and knaves. Suppose A asserts a proposition which is expressed by a formula X. Then the definition of knights leads to the following fact.

- The inhabitant A is a knight *if and only if* X is true.

The fact can be formalized by the formula of bi-implication:

$$A \leftrightarrow X$$

The formalization of the well-formed relation between inhabitants and assertions can be justified by the truth table of bi-implication, following the case analysis on A of either a knight or a knave.

[1] If the answer is "da" then ask C, otherwise (i.e., "ja") ask B.

A	X	$A \leftrightarrow X$
t	t	t
t	f	f
f	t	f
f	f	t

Now recall problem 1.3, and then we obtain the formula $A \leftrightarrow \neg A \wedge \neg B$ from the statement of A. The solution of the puzzle is given by solving the satisfiability problem (SAT) of the formula as follows:

A	B	$\neg A$	$\neg B$	$\neg A \wedge \neg B$	$A \leftrightarrow \neg A \wedge \neg B$
t	t	f	f	f	f
t	f	f	t	f	f
f	t	t	f	f	t
f	f	t	t	t	f

In other words, the puzzle and its solution form the relation of logical consequence:

$$A \leftrightarrow \neg A \wedge \neg B \models \neg A \wedge B$$

This methodology for solving the puzzles of lying and truth-telling can be naturally extended to that for the hardest logic puzzle ever in the next.

2 Formalization

For the extension, the binary relation between inhabitants and assertions should be replaced with a ternary form of questions, answerers, and answers. Firstly, let X be a propositional variable for a question, which means either true or false, respectively represented by 1 or 0. Secondly, let A, B be propositional variables for answerers A, B, which mean either[2] True or False, respectively represented by 1 or 0. Lastly, let Y be a propositional variable for an answer. Here, an answer means either yes or no, respectively represented by 1 or 0. Instead, Y may be used for an answer da-ja, whose meaning is also either 1 or 0, but not fixed yet.

If we ask a question X of an answerer A and obtain an answer Y, then the situation is depicted by the following diagram.

$$X \longrightarrow \boxed{\;A\;} \longrightarrow Y$$

We formalize this relation of question-answerer-answer by the ternary form with the logical connective of bi-implication.

$$X \leftrightarrow A \leftrightarrow Y$$

[2] In the puzzle of Boolos [1], one has Random in addition, but for the formalization here answerers are supposed to be either True or False. According to the solutions [1, 5, 4], Random can be handled by a certain *strategy* of asking questions which can be formalized here in terms of the ternary form.

Note trivial facts that $\leftrightarrow$ is symmetric and associative and that a tautology is the unit. An adequacy of the formalization can be expounded by the truth values (0, 1) semantics. Let **Prop** be the set of propositions (formulae), and $\{0, 1\}$ for the set of truth values. We write v for the assignment $v : \mathbf{Prop} \to \{0, 1\}$. Now the consistent relation of question-answerer-answer is stated as follows:

$$v(X \leftrightarrow A \leftrightarrow Y) = 1,$$

under the assignment v such that $v(A := \text{"True"}) = v(Y := \text{"yes"}) = 1$. This statement can be justified by the case analysis on A, as follows.

- Case A of "True": $v(X \leftrightarrow Y) = 1$

$$X \longrightarrow \boxed{\text{True}} \longrightarrow Y$$

- Case A of "False": $v(X \leftrightarrow Y) = 0$

$$X \longrightarrow \boxed{\text{False}} \longrightarrow Y$$

Let us formalize the embedded question [5, 4]. Recall the first one form the solution:

- Ask god A Q_1: if I asked you if god B was Random, would you say da?

$$\text{“}X \to \boxed{A} \to Y\text{”} \longrightarrow \boxed{A} \longrightarrow ?_1$$

where $X :=$ "B is Random", $Y :=$ "da".

Now suppose that A's answer $?_1$ is "da". Then this situation is formalized by the formula $(X \leftrightarrow A \leftrightarrow Y) \leftrightarrow A \leftrightarrow Y$, and hence for any assignment v we have the following equation

$$v((X \leftrightarrow A \leftrightarrow Y) \leftrightarrow A \leftrightarrow Y) = v(X),$$

since $A \leftrightarrow A$ and $Y \leftrightarrow Y$ are tautologies. This implies that one can conclude $v(X) = 1$. Analogously, we conclude $v(X) = 0$ if the answer $?_1$ is "ja". This is the reason why we can identify the truth value of X from $?_1$, even if we know neither the semantics of A nor that of "da" ("ja").

Moreover, this equation means that the following forms of question-answerer-answer are equivalent to each other under the semantics.

$$\begin{array}{rl} & \text{“}X \to \boxed{A} \to Y\text{”} \longrightarrow \boxed{A} \longrightarrow Y \\ \Longleftrightarrow & X \longrightarrow \boxed{\text{True}} \longrightarrow \text{“yes”} \\ \Longleftrightarrow & \text{“}X \text{ iff } A \text{ iff } Y\text{”} \longrightarrow \boxed{A} \longrightarrow Y \end{array}$$

Next, in order to analyze Boolos' solution in the style of biconditional questions [1], we recall his solution:

- Ask god A Q1: does da means yes iff, you are True iff B is Random?

$$\text{“}(Y \text{ iff } Y_1) \text{ iff, } A \text{ iff } X\text{”} \longrightarrow \boxed{A} \longrightarrow ?$$

where $X :=$ “B is Random”, $Y :=$ “da”, $Y_1 :=$ “yes”.

Here, suppose A's answer ? is “da”. Then we obtain the formula $((Y \leftrightarrow Y_1) \leftrightarrow (A \leftrightarrow X)) \leftrightarrow A \leftrightarrow Y$, so that for any assignment v the equation holds true

$$v(((Y \leftrightarrow Y_1) \leftrightarrow A \leftrightarrow X) \leftrightarrow A \leftrightarrow Y) = v(X \leftrightarrow Y_1).$$

Thanks to the semantics $v(X \leftrightarrow Y_1) = 1$, this form in the style of biconditional questions is equivalent to the definition of “True” as the embedded question is.

$$X \longrightarrow \boxed{\text{True}} \longrightarrow \text{“yes”}$$

Finally, we quote yet another puzzle from the film Labyrinth [6]. There are two doors with two guards, to say, A and B, either True or False. Your task is to determine whether which door leads to the castle by asking A or B one question. What kind of questions makes you reach the castle? In the film, Sarah asked A: Would he (B) tell me that this door leads to the castle? That is,

$$\text{“}X \rightarrow \boxed{B} \rightarrow Y\text{”} \longrightarrow \boxed{A} \longrightarrow ?$$

where $X :=$ “This door leads to the castle”, $Y :=$ “yes”. Now, suppose A's answer ? is “no”, i.e., $\neg Y$. Then we obtain the formula $(X \leftrightarrow B \leftrightarrow Y) \leftrightarrow A \leftrightarrow \neg Y$, and hence for any assignment v the equation holds

$$v((X \leftrightarrow B \leftrightarrow Y) \leftrightarrow A \leftrightarrow \neg Y) = v(X),$$

where $B \leftrightarrow \neg A$. Note also that $A \leftrightarrow \neg A$ and $Y \leftrightarrow \neg Y$ are equivalent to the contradiction $\perp$ where $v(\perp) = 0$ for any v, and of course, $\perp \leftrightarrow \perp$ is a tautology. Moreover, this methodology is still available for the setting of da-ja instead of yes-no.

3 Boolos' puzzle revisited

Following our formalization we summarize the solution [5, 4] which consists of the questions Q_1, Q_2, Q_3 in this order, depending on the answer $?_1$:

1. Q_1 (Ask god A: if I asked you if god B was Random, would you say da?)

$$\text{“B} = \text{R} \rightarrow \boxed{A} \rightarrow \text{da”} \longrightarrow \boxed{A} \longrightarrow ?_1$$

2. $Q_2(Z := C)$ if $?_1 = \text{da}$ ($Q_2(Z := B)$ otherwise (i.e. $?_1 = \text{ja}$))

$$\text{“}Z = \text{T} \rightarrow \boxed{Z} \rightarrow \text{da”} \longrightarrow \boxed{Z} \longrightarrow ?_2$$

3. $Q_3(Z := C)$ (otherwise $Q_3(Z := B)$)

$$\text{“A = R} \to \boxed{Z} \to \text{da”} \longrightarrow \boxed{\quad Z \quad} \longrightarrow ?_3$$

As a solution we have 3! patterns consisting of R, T, or F for $\langle \text{A}, \text{B}, \text{C} \rangle$, and 2^3 patterns $\langle ?_1, ?_2, ?_3 \rangle$ for an answer to $\langle Q_1, Q_2, Q_3 \rangle$. Every candidate for $\langle \text{A}, \text{B}, \text{C} \rangle$ and $\langle Q_1, Q_2, Q_3 \rangle$, and the correlation are compacted in the following table.

	A	B	C	Q_1	$Q_2(C)$	$Q_3(C)$	?
1-1	R	T	F	da	ja	da	1
2-1	R	F	T	da	da	da	2
3	T	R	F	da	ja	ja	3
4	F	R	T	da	da	ja	4
	A	B	C	Q_1	$Q_2(B)$	$Q_3(B)$	?
1-2	R	T	F	ja	da	da	5
2-2	R	F	T	ja	ja	da	6
5	T	F	R	ja	ja	ja	7
6	F	T	R	ja	da	ja	8

4 Remarks and observations

We introduced a simple formalization of puzzles of questions-answerers-answer

$$X \longrightarrow \boxed{\quad A \quad} \longrightarrow Y$$

by using bi-implications $X \leftrightarrow A \leftrightarrow Y$, and the formalization can be justified under the truth values semantics. This method makes it possible to apply algebraic properties of the connective rather than making truth-tables. It turns out that Boolos' solution in the style of biconditional questions [1] and the solution in the style of embedded question [5, 4] are logically equivalent under this semantics. The hardest logic puzzle ever can be regarded as a generalization of Smullyan's Knights-Knaves puzzles [3]. We show that this method is also applicable elegantly to the puzzle in the film Labyrinth [6]. Moreover, this method can formalize naturally n-times nesting of embedded question:

$$\text{“ “}X \to \boxed{A_1} \to Y_1\text{”} \to \boxed{A_2} \to Y_2\text{”} \to \cdots \to \boxed{A_n} \to Y_n$$

We make remarks on the definition of Random. According to [1], 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. Random will answer da or ja when asked any yes-no question, so that the definition can be depicted in the following.

$$X \longrightarrow \boxed{\quad R \quad} \longrightarrow \begin{cases} \text{“da”} \begin{cases} \text{yes} & \text{for } X = 1 \\ \text{no} & \text{for } X = 0 \end{cases} & \text{if } R = \text{T}, \\ \text{“ja”} \begin{cases} \text{yes} & \text{for } X = 0 \\ \text{no} & \text{for } X = 1 \end{cases} & \text{if } R = \text{F}. \end{cases}$$

That is, to put it simply, the diagram becomes the following one.

$$X \longrightarrow \boxed{\quad R \quad} \longrightarrow \begin{cases} \text{“da”} & \text{if heads,} \\ \text{“ja”} & \text{if tails.} \end{cases}$$

Here, we established a very good question (embedded question), so that one can verify whether X is 1 or 0 independent of the values of “da” and R. However, if the question X contains R as an answerer (i.e., embedded questions for R), when and how often does R flip a coin? If Random flips a coin everywhere for each question, for instance, in the diagram below:

$$\text{“}X \to \boxed{R} \to Y\text{”} \longrightarrow \boxed{\quad R \quad} \longrightarrow \ ?,$$

then one cannot enjoy the good property of the embedded questions. Fortunately, the solutions [1, 5, 4] of the puzzle can be provided without answering the question of when-and-how-often, by using an elegant strategy of asking three questions (Q_1, Q_2, Q_3) following the case analysis such that $(\mathrm{A} = \mathrm{R}) \vee (\mathrm{A} \neq \mathrm{R})$ based on the law of the excluded middle.

5 A general form of the puzzle

Finally, we provide a partial solution to the even harder puzzle (2) in [5]:

(2) Suppose the puzzle is as before[3], but one god is Random, and the other two may be either both True, or both False, or one True and one False; is it possible to identify all of the gods in three questions?

Let Z be True (T), False (F), or Random (R). Then the following question

$$\text{“B} = Z \to \boxed{A} \to \text{da”} \longrightarrow \boxed{\quad A \quad} \longrightarrow \ ?$$

is denoted simply by $A : \mathrm{B} = Z$. We employ the binary tree below

$$\frac{\overline{A_1 : \mathrm{B}_1 = Z_1}\ ja \qquad \overline{A_2 : \mathrm{B}_2 = Z_2}\ da}{A : B = Z}\ da|ja$$

to represent that if the answer ? of the question is da then the next question is $A_1 : \mathrm{B}_1 = Z_1$ and the answer of A_1 is ja, and that if the answer ? is ja then the next question is $A_2 : \mathrm{B}_2 = Z_2$ and the answer of A_2 is da. We use this tree representation of sequences of questions and answers. Now a solution to the even harder puzzle (2) can be depicted by the following tree based on the case analysis on $A = R$ or $A \neq R$, starting from $A : B = R$.

Case of $A = R$ where the answer of $A : B = R$ is da:

[3] Of course, this is the hardest logic puzzle ever [1].

$$\dfrac{\dfrac{\dfrac{\overline{C:A=T}\ ja \quad \overline{C:A=T}\ {ja \atop da|ja}}{\dfrac{C:B=T}{C:A=R}\ da} \quad \dfrac{\overline{C:A=T}\ ja \quad \overline{C:A=T}\ {ja \atop da|ja}}{\dfrac{C:B=T}{C:A=R}\ {da \atop da|ja}}}{C:C=T} \qquad \overline{B:C=T}\ da|ja}{A:B=R}$$

Case of $A = R$ where the answer of $A : B = R$ is ja:

$$\dfrac{\overline{C:C=T} \qquad \dfrac{\dfrac{\overline{B:A=T}\ ja \quad \overline{B:A=T}\ {ja \atop da|ja}}{\dfrac{B:B=T}{B:A=R}\ da} \quad \dfrac{\overline{B:A=T}\ ja \quad \overline{B:A=T}\ {ja \atop da|ja}}{\dfrac{B:B=T}{B:A=R}\ {da \atop da|ja}}}{B:C=T}\ da|ja}{A:B=R}$$

By the tree, for instance, the sequence of da, da, da, da, ja in the five questions from $A : B = R$ to $C : A = T$ means that $A = R$, $B = T$, and $C = T$. The sequence of ja, ja, da, ja, ja means that $A = R$, $B = F$, and $C = F$.

Case of $A \neq R$:

$$\dfrac{\dfrac{\dfrac{\dfrac{\overline{C:A=T}\ {da|ja \atop ja}}{C:B=T}\ ja}{C:A=R} \quad \dfrac{\dfrac{\overline{C:A=T}\ {da|ja \atop ja}}{C:B=T}\ ja}{C:A=R}\ da|ja}{C:C=T} \qquad \dfrac{\dfrac{\overline{B:A=T}\ da|ja \quad \overline{B:A=T}\ {da|ja \atop da|ja}}{B:B=T}\ ja}{\dfrac{B:A=R}{B:C=T}\ {ja \atop da|ja}}}{A:B=R}$$

By the tree, the sequence da, da, ja, ja, da in the five questions means that $A = T$, $B = R$, and $C = T$; and ja, ja, ja, ja, ja means that $A = F$, $B = F$, and $C = R$.

At the end, we pose a general form of the puzzle. Let $G_n = \{A_1, A_2, \ldots, A_n\}$ $(n \geq 1)$ be the set of gods where A_i $(1 \leq i \leq n)$ is Random, True, or False. Let $|G_n|_R$ be the number of Random in G_n, and suppose $|G_n|_R < n$. Then the following fundamental puzzle is suggested.

1. Is it possible to identify the non-Random god in G_n?

2. In particular, is it possible to identify the non-Random god for the case where $|G_2|_R = 1$?

Remarked that the even harder puzzle (1) in [5] is now an instance of $|G_3|_R = 2$. We have already provided a solution in five questions to one case where $|G_3|_R = 1$. We conjecture that it is impossible to identify the non-Random god for the case where $|G_2|_R = 1$ by using a finite sequence of questions.

References

[1] George Boolos: The Hardest Logic Puzzle Ever, The Harvard Review of Philosophy **6**, pp. 62–65, 1996.

[2] K. Fujita: On formalization of logic puzzles *á la* Smullyan, Kyoto University RIMS Kôkyûroku No. 2193 (Logic, Language, Algebraic system and Related Areas in Computer Science), pp. 18–24, 2021.

[3] Raymond Smullyan: Logical Labyrinths, A K Peters, Ltd 2009. (高橋昌一郎 監訳，川辺治之 訳：スマリヤン 記号論理学，丸善出版 2013)

[4] Brian Rabern and Landon Rabern: A simple solution to the hardest logic puzzle ever, ANALYSIS **68.2**, pp 105–112, 2008.

[5] Tim S. Roberts: Some Thoughts About The Hardest Logic Puzzle Ever, Journal of Philosophical Logic **30**, pp. 609–612, 2001.

[6] Jim Henson with George Lucas: Labyrinth, Lucasfilm Ltd. 1986.

Ken-etsu Fujita

Gunma University, JAPAN

E-mail address: fujita@gunma-u.ac.jp

Toshihiko Kurata

Hosei University, JAPAN

E-mail address: kurata@hosei.ac.jp