Knights and Knaves

Knights and Knaves is a type of logic puzzle where some characters can only answer questions truthfully, and others only falsely. The name was coined by Raymond Smullyan in his 1978 work *What Is the Name of This Book?*.[1]

The puzzles are set on a fictional island where all inhabitants are either knights, who always tell the truth, or knaves, who always lie. The puzzles involve a visitor to the island who meets small groups of inhabitants. Usually the aim is for the visitor to deduce the inhabitants' type from their statements, but some puzzles of this type ask for other facts to be deduced. The puzzle may also be to determine a yes-no question which the visitor can ask in order to discover a particular piece of information.

One of Smullyan's examples of this type of puzzle involves three inhabitants referred to as A, B and C. The visitor asks A what type he is, but does not hear A's answer. B then says "A said that he is a knave" and C says "Don't believe B; he is lying!".[2] To solve the puzzle, note that no inhabitant can say that he is a knave. Therefore, B's statement must be untrue, so he is a knave, making C's statement true, so he is a knight. Since A's answer invariably would be "I'm a knight", it is not possible to determine whether A is a knight or knave from the information provided.

Maurice Kraitchik presents the same puzzle in the 1953 book *Mathematical Recreations*, where two groups on a remote island – the Arbus and the Bosnins – either lie or tell the truth, and respond to the same question as above.[3]

In some variations, inhabitants may also be alternators, who alternate between lying and telling the truth, or normals, who can say whatever they want.[2] A further complication is that the inhabitants may answer yes/no questions in their own language, and the visitor knows that "bal" and "da" mean "yes" and "no" but does not know which is which. These types of puzzles were a major inspiration for what has become known as "the hardest logic puzzle ever".

Contents

Examples

A large class of elementary logical puzzles can be solved using the laws of Boolean algebra and logic truth tables. Familiarity with boolean algebra and its simplification process will help with understanding the following examples.

John and Bill are residents of the island of knights and knaves.

Both knaves

John says "We are both knaves."

In this case, John is a knave and Bill is a knight. John's statement cannot be true because a knave admitting to being a knave would b the same as a liar telling the lie "I am a liar", which is known as the liar paradox. Since John is a knave this means he must have been lying about them *both* being knaves, and so Bill is a knight.

Same or different kinds

John says "We are the same kind.", but Bill says "We are of different kinds."

In this scenario they are making contradictory statements and so one must be a knight and one must be a knave. Since that is exactly what Bill said, Bill must be the knight, and John is the knave.

Identity alone

If all we want to know is whether a man is a knight or a knave, this can be tested by simply asking a question to which the answer is already known. In the film *The Enigma of Kaspar Hauser*, Kaspar solves the puzzle of whether a man is a knight or a knave by suggesting asking the man "whether he was a tree frog".

Fork in the road

"John and Bill are standing at a fork in the road. John is standing in front of the left road, and Bill is standing in front of the right road. One of them is a knight and the other a knave, but you don't know which. You also know that one road leads to Death, and the other leads to Freedom. By asking one yes–no question, can you determine the road to Freedom?"

This is perhaps the most famous rendition of this type of puzzle. This version of the puzzle was further popularised by a scene in the 1986 fantasy film, *Labyrinth*, in which the protagonist finds herself faced with two doors with guardians who follow the rules of the puzzle. One door leads to the castle at the centre of the labyrinth, and one to certain death.

There are several ways to find out which way leads to freedom. All can be determined by using Boolean algebra and a truth table.

In *Labyrinth*, the protagonist's solution is to ask: "Answer yes or no. Would he tell me that this door leads to the castle?". As the question is about the other persons response, it will either result in a lie about the other truthfully saying which door leads to freedom or a truth about the other lying about which door leads to freedom. Either way, the response will be the opposite of what a truthful person would say had you asked them whether their door leads to freedom.

A simpler solution is to ask either man if they themselves would tell you that their path leads to freedom. For example, in the case where the knight's door in fact leads to freedom, the knight would answer "yes" (as he would tell you that his own path led to freedom) and the knave "no" (although he would also tell you that his own path led to freedom, he would lie about whether he would tell you this).

References

1. George Boolos, John P. Burgess, Richard C. Jeffrey, *Logic, logic, and logic* (Harvard University Press, 1999).
2. Smullyan, Raymond (1978). *What is the Name of this Book?* Prentice-Hall.
3. Kraitchik, Maurice (1953). *Mathematical Recreations* Dover. ISBN 0486201635.

External links

- A note on some philosophical implications of the Knights and Knaves puzzle for the concept of knowability
- A complete list and analysis of Knight, Knave, and Spy puzzles, where spies are able to lie or tell the truth.