Knights and Knaves

Peter Rowlett

Sheffield Hallam University

p.rowlett@shu.ac.uk

Smullyan's island

- Raymond Smullyan sets a series of puzzles on an island in which there are two types of inhabitants:
  - 'knights', who always tell the truth;
  - 'knaves', who always lie.
- Everyone on the island is either a knight or a knave.

Example

You are a visitor to the island and meet three of the inhabitants: Alice, Bob, and Carol. You ask Alice "How many knights are among you?" Alice answers, but mumbles so you don't hear her answer. You ask Bob, "What did Alice say?" Bob replies "Alice said there is one knight among us." At this point Carol said "Bob is lying."

What types are Bob and Carol?

Propositions

A proposition is a sentence that has a truth value, it is unambiguously true or false. Which of the following are propositions?

1. "One plus one equals two."
2. "Come here."
3. "The moon is made of green cheese."
4. "Would you like a cup of tea?"
5. "If $f(x) = x^2$, then $f(4) = 18$."
6. "5=5".

Propositions

Let's set out two propositions about inhabitants of Smullyan's island:

- p: "Alice is a knight."
- q: "Bob is a knight."

Connectives

NOT

- p: "Alice is a knight."
- q: "Bob is a knight."

- NOT is a connective that negates a statement. If p is true, then NOT p is false, and vice versa. We will write NOT p as $\neg p$.
- With our propositions about inhabitants of Smullyan's island, what does $\neg p$ mean?

NOT

- p: "Alice is a knight."
- q: "Bob is a knight."

- NOT is a connective that negates a statement. If p is true, then NOT p is false, and vice versa. We will write NOT p as $\neg p$.
- With our propositions about inhabitants of Smullyan's island, what does $\neg p$ mean?
- Since everyone on Smullyan's island is either a knight or a knave, $\neg p$ is the proposition "Alice is a knave".

NOT

We can represent this information in an arrangement called a truth table.

p	$\neg p$
true	false
false	true

AND

- p: "Alice is a knight."
- q: "Bob is a knight."

- We can combine two propositions using AND, written as $\wedge$. This is only true if both p and q are true.
- With our propositions about inhabitants of Smullyan's island, what does $p \wedge q$ mean?

AND

- p: "Alice is a knight."
- q: "Bob is a knight."

- We can combine two propositions using AND, written as $\wedge$. This is only true if both p and q are true.
- With our propositions about inhabitants of Smullyan's island, what does $p \wedge q$ mean?
- $p \wedge q$ is the proposition "Alice and Bob are both knights."

OR

- p: "Alice is a knight."
- q: "Bob is a knight."

- Another way to combine two propositions is using OR, written $\vee$. This is true if at least one of p and q are true.
- With our propositions about inhabitants of Smullyan's island, what does $p \vee q$ mean?

OR

- p: "Alice is a knight."
- q: "Bob is a knight."

- Another way to combine two propositions is using OR, written $\vee$. This is true if at least one of p and q are true.
- With our propositions about inhabitants of Smullyan's island, what does $p \vee q$ mean?
- $p \vee q$ is the proposition "Alice is a knight, or Bob is a knight, or both Alice and Bob are knights."

XOR

- p: "Alice is a knight."
- q: "Bob is a knight."

- The exclusive OR (XOR), written $\oplus$ is used when either p or q are true but not both.
- With our propositions about inhabitants of Smullyan's island, what does $p \oplus q$ mean?

XOR

- p: "Alice is a knight."
- q: "Bob is a knight."

- The exclusive OR (XOR), written $\oplus$ is used when either p or q are true but not both.
- With our propositions about inhabitants of Smullyan's island, what does $p \oplus q$ mean?
- $p \oplus q$ is the proposition "Alice is a knight, or Bob is a knight, but not both", or equivalently, "Either Alice is a knight and Bob is a knave, or Alice is a knave and Bob is a knight".

Truth tables for AND, OR, and XOR

p	q	$p \wedge q$	$p \vee q$	$p \oplus q$
true	true	true	true	false
true	false	false	true	true
false	true	false	true	true
false	false	false	false	false

Truth tables

- We can use truth tables to organise the information in a puzzle and help us solve it. For example, consider this puzzle.

Which of these statements, if any, are true?

1. In this list exactly one statement is false.
2. In this list exactly two statements are false.
3. In this list exactly three statements are false.
4. In this list exactly four statements are false.
5. In this list exactly five statements are false.

- Let's suppose statement 1 is true. There are then four options for the truth of the other statements.

1	2	3	4	5
true	false	true	true	true
true	true	false	true	true
true	true	true	false	true
true	true	true	true	false

- In each of these we can find a contradiction. For example, in the first row, statement 3 is true, which says exactly 3 statements are false. Since we only have one false statement, this cannot be.

Truth tables

- Use truth tables to solve this puzzle.

Which of these statements, if any, are true?

1. In this list exactly one statement is false.
2. In this list exactly two statements are false.
3. In this list exactly three statements are false.
4. In this list exactly four statements are false.
5. In this list exactly five statements are false.

Answer

1	2	3	4	5
false	false	false	true	false

The answer to the puzzle is that statement 4 is the only one that is true.

Generalise

- This puzzle form 'exactly k statements on this list are false' for $k = 1, \ldots, n$ extends to any $n > 2$, with the answer $n - 1$. It is often presented in the $n = 10$ case.
- The $n = 1$ case is interesting.

Which of these statements, if any, are true?

1. In this list exactly one statement is false.

Generalise

- This puzzle form 'exactly k statements on this list are false' for $k = 1, \ldots, n$ extends to any $n > 2$, with the answer $n - 1$. It is often presented in the $n = 10$ case.
- The $n = 1$ case is interesting.

Which of these statements, if any, are true?

1. In this list exactly one statement is false.

- This arrangement is known as the liar's paradox, often presented as 'this sentence is false'.