A Method of Solving Knights-And-Knaves Questions

There is an island far off in the Pacific, called the Island of Knights and Knaves. On this island, there are people called knights, who always tell the truth, and people called knaves, who always lie. The two types are indistinguishable by sight.

1. You meet two people, A and B. A says: “I’m a knave but B isn’t.” What are A and B?
2. Suppose A says: “If I am a knight, then so is B.” Can it be determined what A and B are?
3. Suppose you know that A and B are either both knights or both knaves. What do you make of A’s statement “If B is a knight, then I am a knave”?
4. Suppose A says: “We are both knights” and B says “Either A is a knight or I am a knight, but not both.” What can you conclude?

Step-by-step solutions

We describe a method for solving such puzzles below. Note that many other approaches can be taken, for example, a case-by-case analysis, which we shall examine later in the course.

A word of caution: Each person’s statement should be considered as a compound proposition, and its truth value examined as such. For example, if a knight says “p or q”, then \( p \lor q \) is true, but \( p \) itself can be false.

1. **STEP I:** Define the following propositional variables:
   - \( p \): “A is a knight.”
   - \( q \): “B is a knight.”

   **STEP II:** Write any statements made by the Islanders as compound propositions in \( p \) and \( q \):
   - A said: \( \neg p \land q \).

   **STEP III:** Observe that, since knights always speak the truth and knaves always lie, the truth value of \( p \) must coincide with the truth value of any statement made by A; similarly, the truth value of \( q \) must coincide with the truth value of any statement made by B.

   In our case, we conclude that \( p \) and \( \neg p \land q \) must have the same truth value.

   **STEP IV:** Construct a truth table to determine which truth assignments for \( p \) and \( q \) (if any) will make the truth value of \( p \) equal to the truth value of any statement made by A, and the truth value of \( q \) equal to the truth value of any statement made by B.

   In our case, we construct a truth table for \( \neg p \land q \) and compare its truth value to that of \( p \):

<table>
<thead>
<tr>
<th>( p )</th>
<th>( q )</th>
<th>( \neg p \land q )</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>F</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
</tbody>
</table>

   The truth values of \( p \) and \( \neg p \land q \) coincide only in the case when \( p \) and \( q \) are both false. Hence A and B must both be knaves.
2. **STEP I:** Define $p$ and $q$ as above.

   **STEP II:** A said: $p \rightarrow q$.

   **STEP III:** Observe that $p$ and $p \rightarrow q$ must have the same truth value.

   **STEP IV:** Construct a truth table for $p \rightarrow q$ and compare its truth value to that of $p$:

   $\begin{array}{ccc}
   p & q & p \rightarrow q \\
   T & T & T \\
   T & F & F \\
   F & T & T \\
   F & F & T \\
   \end{array}$

   The truth values of $p$ and $p \rightarrow q$ coincide only in the case when $p$ and $q$ are both true. Hence A and B must both be knights.

3. **STEP I:** Define $p$ and $q$ as above.

   **STEP II:** A said: $q \rightarrow \neg p$.

   **STEP III:** Observe that $p$ and $q \rightarrow \neg p$ must have the same truth value.

   **STEP IV:** Construct a truth table for $q \rightarrow \neg p$ and compare its truth value to that of $p$. Note that since A and B are either both knights or both knaves, only the following two rows are needed:

   $\begin{array}{ccc}
   p & q & q \rightarrow \neg p \\
   T & T & F \\
   F & T & T \\
   \end{array}$

   Observe that the truth values of $p$ and $q \rightarrow \neg p$ do not agree in either of the two cases. We conclude that (with the assumption that A and B are either both knights or both knaves) no Islander would make such a claim.

4. **STEP I:** Define $p$ and $q$ as above.

   **STEP II:** A said $p \land q$, and B said $p \oplus q$ (or, equivalently, $\neg(p \leftrightarrow q)$).

   **STEP III:** Observe that $p$ and $p \land q$ must have the same truth value, as well as $q$ and $p \oplus q$.

   **STEP IV:** Construct a truth table for $p \land q$ and $p \oplus q$. Compare the truth values of $p$ and $p \land q$, as well as $q$ and $p \oplus q$.

   $\begin{array}{cccc}
   p & q & p \land q & p \oplus q \\
   T & T & T & F \\
   T & F & F & T \\
   F & T & F & T \\
   F & F & F & F \\
   \end{array}$

   Observe that the truth values of $p$ and $p \land q$ agree in rows 1, 3, and 4, and the truth values of $q$ and $p \oplus q$ agree in rows 3 and 4. Hence both pairs agree in rows 3 and 4, that is, when $p$ is false and $q$ is either true or false. We conclude that A must be a knave, while the identity of person B can not be determined.