Smullyan's Knights and Knaves

The logician Raymond Smullyan popularized logic puzzles set on the island of knights and knaves. The island has two kinds of inhabitants: knights, who always tell the truth, and knaves, who always lie. In a typical puzzle, an observer encounters a group of inhabitants, hears them make a collection of statements, and then tries to work out which of the inhabitants in the group are knights and which are knaves. Sometimes the objective is to find out something about the environment (e.g., which road actually leads to the airport.)

Here is a simple example puzzle. We encounter two inhabitants, conveniently named A and B. They make the following statements:

```
A: We are both knights.
B: No, we aren't!
```

From this we are supposed to derive something about the types of A and B, that is, whether each is a knight or a knave.

One way to go about this is to notice that they cannot both be telling the truth, because they contradict each other. Thus, A's statement cannot be true (because if they are both knights, they must both tell the truth.) This allows us to conclude that A must be a knave, because only a knave can say false things. And what about B? B's statement is true, because they are certainly not both knights, and since only a knight can make a true statement, B must be a knight. So the (unique) solution to this puzzle is that A is a knave and B is a knight.

We consider a more complex puzzle, with statements by three inhabitants: A, B, C.

```
A: Exactly one of us is a knave.
B: At least one of us is a knight.
C: None of us is a knave.
```

Let us consider C's statement first. If it is true, then all of A, B and C must be knights, so all their statements must be true. But A's statement claims that exactly one of A, B, and C is a knave, which is false if they are all knights. This would entail a knight saying something false, so C's statement cannot be true. Thus, we learn that C is a knave.

Now we consider A's statement. If it is true, then there is exactly one knave (namely, C), so A and B must be knights. This is a consistent situation, since A's statement is true, and B's statement is true, and C's statement is false in this case. Hence one possible scenario is that A and B are knights, and C is a knave. But there may be other possible scenarios.

Suppose A's statement is false, and therefore A is a knave. Then there must be 0, 2, or 3 knaves in the group (that is, not exactly one.) There cannot be no knaves in the group (because C is known to be a knave, and in the case we are considering, A is also a knave.)

There could be two knaves if B is a knight. This is possible because then what B says is true. So another possibility is that A is a knave, B is a knight, and C is a knave.

There could be three knaves in the group if B is a knave. This also is possible because then what B says is false (because there are no knights in the group in this case.) Thus the final possibility is that all three of A, B, and C are knaves.

Some of the puzzles give statements that could not have been made under the assumptions about knights and knaves. For example, we are not going to come upon an inhabitant A who says:

```
A: I am a knave.
```

Why not? There are further elaborations of these puzzles, including normals and sane and insane humans and vampires – see Smullyan's books for these and many other types of logic puzzles.