**Definition** A *knight-knave island* is an island in which each native is classified as either a *knight* or a *knave*. Knights make only true statements and knaves make only false ones.

Our two main characters are a logician L who visits an island and meets a native N. Let \( k \) be the proposition that N is a knight. Whenever N asserts a proposition \( p \), the reality of the situation is that \( k \equiv p \) is true (N is a knight if and only if \( p \)). For any proposition \( p \), let \( Bp \) be the proposition that L does or will believe \( p \) — i.e., L does or will decide that \( p \) is true — the alternate notation \( B : p \) will sometimes be used to denote \( Bp \).

**Definition** A *proof* is a finite list whose elements are axioms or results of applying inference rules to previous elements in the list. The last element in the list is what has been proved.

Note that elements in the list comprising a proof are assumed — they are *axioms* — or they are justified by applying some inference rule to previously proven statements. Moreover, truncating the list by discarding statements after any element in a proof yields a proof of that element.

Inference rules include

\[
\begin{align*}
\text{\( p \) is a tautology} & \quad B : \quad p \\
\text{\( B : \quad p \Rightarrow q \)} & \quad B : \quad p \\
\text{\( k = \text{N is a knight} \)} & \quad B : \quad q \\
\text{\( \text{N asserts } p \)} & \quad B : \quad k \equiv p
\end{align*}
\]

The definition above makes it problematic to accept as proof explanations like the following:

The logician believes he is with a knave, because the native said it was a sunny day during a thunderstorm.

Initial attempts to fit that explanation into the framework of proof might produce a list like

- There was a thunderstorm.
- The native said it was a sunny day.
- The native was not telling the truth.
- The logician believes the native is a knave.

Difficulties arise when classifying elements as either axioms or results of inference rules; what inference rule justifies the last element? The first inference rule does not apply unless “the native is a knave” is a tautology, but a tautology is a propositional expression which always evaluates to true; must \( \neg k \) always be true? The second inference rule does not apply since that would require previous elements in the list to address what the logician believes (but there are none). The last inference rule does not apply since it has a result concerning the concept of *knight* (i.e., \( k \)), not the concept of *knave* (i.e., \( \neg k \)).
If the list above is to be a proof, the only alternative is to classify the last element – which is the statement being proved – as an axiom. In that case the proof could be shortened to

- The logician believes the native is a knave.

In other words, we would not be coming to a provable conclusion about the logician based on evidence, we would instead be deciding what he believes based on assuming what he believes!

Let us try again to fit the explanation into the framework of proof. To streamline exposition and facilitate unambiguous discussion, we introduce names for key propositions. Consider the following:

1. \( p = \) it was a sunny day
2. \( k = N \) is a knight
3. \( N \) asserts \( p \)
4. \( B : k \equiv p \)
5. \( t = \) there was a thunderstorm
6. \( t \)
7. \( t \Rightarrow \neg p \)
8. \( \neg p \)
9. \( B \neg k \)

Elements 1, 2, 5 are reasonable axioms; we are free to assume statements can be represented by names. Elements 3, 6 may be classified as axioms; we are free to assume information that is given. Element 4 is the result of applying the third inference rule (see previous page) to elements 2, 3.

So far, so good, but our proof attempt looks questionable beginning at element 7. Whereas we are free to assume any tautology, the proposition \( t \Rightarrow \neg p \) is not always true:

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

Even if we were to accept item 7 as an axiom — within the context of the specific semantics given to \( t \) and \( p \) by items 5 and 1 it may be reasonable to assume that a thunderstorm implies it was not a sunny day — how could we justify line 8? For it to follow from lines 6 and 7 would seem to require an inference rule like

\[
\begin{align*}
p & \Rightarrow q \\
p & \Rightarrow q \\
q & 
\end{align*}
\]

However, we have not (yet) been given that inference rule!

Better than two questionable assumptions — the first concerning \( t \Rightarrow \neg p \) and the second concerning an inference rule — is to make only one, namely, assume that \( \neg p \) is given. In that case, our proof attempt would shorten to
1. $p = \text{ it was a sunny day}$
2. $k = N$ is a knight
3. $N$ asserts $p$
4. $B : k \equiv p$
5. $\neg p$
6. $B \neg k$

All would be fine, except for item 6 which may be impossible to prove. It is dubious that $\neg p$ would necessarily shed light on $B \neg k$. The fact that $\neg p$ is true may well be irrelevant — to the logician — if the logician does not happen to believe it!

To make progress, we might assume $B \neg p$ is given. It should be appreciated that we have lost connection with the initial story line. Taking $B \neg p$ as given based on the story line “...during a thunderstorm...” is unjustified in the absence of a connection between the thunderstorm and what the logician believes!

If we make the assumption $B \neg p$ about what the logician believes, we are in effect revising the initial story line to the following:

The logician believes he is with a knave, because the native said it was a sunny day when the logician believed it wasn’t.

This revised explanation fits easily into the framework of proof as follows:

1. $p = \text{ it was a sunny day}$ \hspace{1cm} (define $p$)
2. $k = N$ is a knight \hspace{1cm} (define $k$)
3. $N$ asserts $p$ \hspace{1cm} (given)
4. $B : k \equiv p$ \hspace{1cm} (third inference rule applied to 2 and 3)
5. $B \neg p$ \hspace{1cm} (given)
6. $(k \equiv p) \Rightarrow (\neg p \Rightarrow \neg k)$ \hspace{1cm} .
7. $B : (k \equiv p) \Rightarrow (\neg p \Rightarrow \neg k)$ \hspace{1cm} .
8. $B : \neg p \Rightarrow \neg k$ \hspace{1cm} .
9. $B \neg k$ \hspace{1cm} .

It is interesting to realize that the native might actually be a knight even though the logician thinks otherwise; the revised story line allows the possibility that the logician is mistaken to believe it wasn’t a sunny day!

**Homework**: Justify items 6 through 9 in the proof above. You may:

- assume statements can be represented by names (i.e., names may be defined),
- assume information that is given,
- assume any tautology (but give a separate demonstration that it is a tautology),
- use the given inference rules (the three “belief schemes” on page one).
Inference rules

Inference rules are typically intended to be truth-preserving. A rule of the form

\[
\begin{align*}
\alpha_1 \\
\vdots \\
\alpha_n \\
\hline 
\beta
\end{align*}
\]

is truth-preserving iff the associated implication "\((\alpha_1 \land \cdots \land \alpha_n) \Rightarrow \beta\)" is a tautology. You may use any truth-preserving inference rule, provided you show the associated implication is a tautology (by way of a truth table, for instance).

Some inference rules need not be truth-preserving. The three given on page one – concerning what the logician believes – are inference rules of that sort; their purpose is to model the logician’s beliefs and patterns of reasoning. Because such inference rules have associated implications that need not be tautologies, they must be either assumed (i.e., explicitly given) or else proved on the basis of what has already been given. To prove the inference rule

\[
\begin{align*}
\alpha_1 \\
\vdots \\
\alpha_n \\
\hline 
\beta
\end{align*}
\]

is to exhibit a proof of \(\beta\) using given axioms, given or previously established inference rules, and \(\alpha_1, \ldots, \alpha_n\) as assumptions.

For example, to say a logician is accurate is to be given the following belief scheme (we call an inference rule relating to what the logician believes a “belief scheme”):

\[
\neg p \\
\hline 
\neg Bp
\]

We prove that the following is a valid belief scheme for an accurate logician

\[
\begin{align*}
Bp \\
\hline 
p
\end{align*}
\]

by way of proof by contradiction, which allows one to use the negation of what is to be proved as an axiom:

1. \(Bp\) given
2. \(\neg p\) assumption (begin proof by contradiction)
3. \(\neg Bp\) the logician is accurate
4. \(Bp \land \neg Bp\) \((\alpha_1 \land \alpha_3) \Rightarrow (\alpha_1 \land \alpha_3)\) is a tautology
5. \((Bp \land \neg Bp) \Rightarrow p\) tautology
6. \(p\) \(((\alpha_4 \Rightarrow \alpha_6) \land \alpha_4) \Rightarrow \alpha_6\) is a tautology (end proof by contradiction)
Item 2 (above) assumes the negation of what is to be proved; in particular, it must be the negation of the last element of the list constituting the proof. Item 3 results from applying the belief scheme for an accurate logician to item 2. Item 4 results from applying the following inference rule to items 1 and 3:

\[
\begin{align*}
\alpha_1 \\
\alpha_3 \\
\hline
\alpha_1 \land \alpha_3
\end{align*}
\]

Item 6 results from applying the following inference rule to items 4 and 5:

\[
\begin{align*}
\alpha_4 \\
\hline
\alpha_4 \Rightarrow \alpha_6
\end{align*}
\]

**Homework:** Suppose the logician is accurate, and \( h \) is such that \( h \equiv B \neg h \). It follows that \( \neg h, \neg Bh, \neg B \neg h \). Completely justify the following proofs:

1. \( h \equiv B \neg h \)
2. \( h \)
3. \( B \neg h \)
4. \( \neg h \)

1. \( h \equiv B \neg h \)
2. \( Bh \)
3. \( h \)
4. \( B \neg h \)
5. \( \neg h \)
6. \( h \land \neg h \)
7. \( (h \land \neg h) \Rightarrow \neg Bh \)
8. \( \neg Bh \)

1. \( h \equiv B \neg h \)
2. \( B \neg h \)
3. \( h \)
4. \( \neg h \)
5. \( h \land \neg h \)
6. \( (h \land \neg h) \Rightarrow \neg B \neg h \)
7. \( \neg B \neg h \)
Conditional proof

Suppose we would like to prove an implication $p \Rightarrow q$ in the context of the given assumptions. We are free to assume $p$ as an axiom — even though it might not be a given assumption — but only during the proof of $q$. Moreover, the section of the proof from the line containing $p$ to the line containing $q$ is indented, and the line immediately following that indented section is the implication $p \Rightarrow q$. The following example illustrates conditional proof.

1. $r$ given
2. $\neg r \lor \neg p \lor q$ given
3. $p$ assumption (begin conditional proof)
4. $\neg p \lor q$ $(\alpha_1 \land (\neg \alpha_1 \lor \alpha_4)) \Rightarrow \alpha_4$ is a tautology
5. $q$ $(\alpha_3 \land (\neg \alpha_3 \lor \alpha_5)) \Rightarrow \alpha_5$ is a tautology
6. $p \Rightarrow q$ (end conditional proof)

**Homework:** Completely justify the proof above (indicate to which items in the proof above the tautologies are applied).