Solutions to problems in Smullyan’s *Logical Labyrinths*

Samuel Fisher

March 7, 2020

Part I – Be Wise, Generalize!

1. The Logic of Lying and Truth-Telling

**Exercise 1.1 (A Classic Case).** *On the day of his arrival, Abercrombie came across three inhabitants, whom we will call A, B and C. He asked A: “Are you a knight or a knave?” A answered, but so indistinctly that Abercrombie could not understand what he said. He then asked B: “What did he say?” B replied: “He said that he is a knave.” At this point, C piped up and said: “Dont believe that; its a lie!”

Was C a knight or a knave?

Note that A cannot have said “I am a knave,” because if A was a knave, then A would have lied, and if A was a knave, then A would have told the truth. Both cases are contradictions. Therefore B must be lying, and therefore C is telling the truth. Hence, C is a knight.

**Problem 1.2 (A Variant).** According to another version of the story, Abercrombie didn’t ask A whether he was a knight or a knave (because he would have known in advance what answer he would get), but instead asked A how many of the three were knaves. Again A answered indistinctly, so Abercrombie asked B what A had said. B then said that A had said that exactly two of them were knaves. Then, as before, C claimed that B was lying.

Is it now possible to determine whether C is a knight or a knave?

Suppose that C is a knave. Then B must be a knight, which means A said there are two knaves. This statement cannot be true, for if it were, it would imply A is a knave, a contradiction. Hence, the statement must be false, implying that A is a knave. But then there are two knaves, namely A and C, so the statement is true, another contradiction. We conclude that C must be a knight.

**Problem 1.3.** Next, Abercrombie met just two inhabitants, A and B. A made the following statement: “Both of us are knaves.” What is A and what is B?

The statement cannot be true, because if it were it would imply that A is both a knight and a knave. Hence, it is false, so A is a knave and B is a knight.

**Problem 1.4.** According to another version of the story, A didn’t say “Both of us are knaves.” All he said was “At least one of us is a knave.”

If this version is correct, what are A and B?

If A is a knave, then the statement is true, a contradiction. So A must be a knight. Then, for the statement to be true, B must be a knave.

**Problem 1.5.** According to still another version, what A actually said was “We are of the same type—that is, we are either both knights or both knaves.”

If this version is correct, then what can be deduced about A and B?

We know that B must be a knight. For if A is a knight, then A and B are of the same type, so B is a knight. If A is a knave, then A and B are of different type, which means that B must be a knight. Both cases mentioned above are logically consistent, so we cannot deduce what A is.

**Problem 1.6.** On one occasion, Abercrombie came across two natives who were lazily lying in the sun.
He asked one of them whether the other one was a knight and got an answer (yes or no). He then asked the other native whether the first one was a knight, and got an answer (yes or no). Were the two answers necessarily the same?

If both A and B are knights, then the answers will both be “yes”. If they are both knaves, then both answers will be “no.” Suppose A is a knave, and B is a knight, then both answers will be “no.” Similarly, if A is a knight and B is a knave, both answers will be “no.”

**Problem 1.7.** On another occasion, Abercrombie came across just one native who was lazily lying in the sun. Abercrombie asked the native his name, and the native replied: “John.” Was the native a knight or a knave?

The only useful piece of information here is that the native is lying in the sun, therefore the native is a knave.

**Problem 1.8.** On another occasion, Abercrombie came across a native and remembered that his name was either Paul or Saul, but couldn’t remember which. He asked him his name, and the native replied “Saul.”

From this, it is not possible to tell whether the native was a knight or a knave, but one can tell with very high probability! How? (This is a genuine problem, not a monkey trick!)

Well, if the native was a knave, they could have picked any name to lie to you with. The odds that they picked one of the two names Abercrombie was thinking of are very low. So, with high probability, the native is a knight.

**Problem 1.9.** In the next incident, Abercrombie came across three natives, A, B, and C, who made the following statements:

A: Exactly one of us is a knave
B: Exactly two of us are knaves
C: All of us are knaves

What type is each?

Clearly at least two of them are lying, so either B or C is telling the truth. But C cannot be telling the truth, because that would imply C is a knave, a contradiction. So B is a knight, and A and C are knaves.

**Problem 1.10 (Who Is The Chief?).** Abercrombie knew that the island had a chief and was curious to find him. He finally narrowed his search down to two brothers named Og and Bog, and knew that one of the two was the chief, but didn’t know which one until they made the following statements:

Og: Bog is the chief and he is a knave!
Bog: Og is not the chief, but he is a knight.

Which one is the chief?

Note that Bog cannot by a knight, because then Og would be too, and Og claims that Bog is a knave. So we know that Bog is a knave, which implies that either Og is chief, or he is a knave (or both). If Og is a knave, then either Bog is not chief or he is a knight. We know that Bog is not a knight, so he cannot be chief. Hence, Og is the chief.

**Problem 1.11.** (Introducing the Nelson Goodman Principle) Suppose that you visit the Island of Knights and Knaves because you have heard a rumour that there is gold buried there. You meet a native and you wish to find out from him whether there really is gold there, but you don’t know whether he is a knight or a knave. You are allowed to ask him only one question answerable by yes or no.

What question would you ask? (The answer involves an important principle discovered by the philosopher Nelson Goodman.)

I would ask “If I asked a native who is not of your type ‘Is there gold buried here,’ would they say yes or no?” If there is gold on the island, the response will be no. If there is no gold on the island, the response will be yes, regardless of whether the native is a knight or a knave. This problem is a variation of the classic problem where you have two guards in front of two doors, one leads to heaven the other to hell. You can ask one question to either of them to figure out which is which.
Problem 1.12 (A Neat Variant). (To be read after the solution of Problem 1.11.) There is an old problem in which it is said that the knights all live in one village and the knaves all live in another. You are standing at a fork in the road and one road leads to the village of knights, and the other to the village of knaves. You wish to go to the village of knights, but you dont know which road is the right one. A native is standing at the fork and you may ask him only one yes/no question. Of course, you could use the Nelson Goodman principle and ask: Are you a knight if and only if the left road leads to the village of knights? There is a much simpler and more natural-sounding question, however, that would do the job – a question using only eight words. Can you find such a question?

I would ask “Does the left road lead to your village?” If the left road leads to the village of the knights, both a knight and knave would respond yes. Otherwise, knights and knaves both answer no.

Problem 1.13 (Three Brothers). Here is another variant using a natural-sounding question: Three triplet brothers on this island are named Larry, Leon, and Tim. They are indistinguishable in appearance, but Larry and Leon, whose names begin with “L,” always lie and hence are knaves, whereas Tim, whose name begins with “T,” is always truthful and hence is a knight.

One day you meet one of the three on the street and wish to know whether or not he is Larry, because Larry owes you money. You are allowed to ask him only one yes/no question, but to prevent you from using the Nelson Goodman principle, your question may not have more than three words! What question would work?

I would ask “Are you Leon?” Only Larry would answer yes to this question.

Problem 1.14 (Two Brothers) Arthur and Robert are twin brothers, indistinguishable in appearance. One of the two always lies and the other always tells the truth, but you are not told whether it is Arthur or Robert who is the liar. You meet one of the two one day and wish to find out whether he is Arthur or Robert. Again you may ask him only one yes/no question, and the question may not contain more than three words. What question would you ask?

I would ask “Does Arthur lie?” If I am speaking to Arthur, I will invariably be told “no.” If I am speaking to Robert, I will hear “yes.”

Problem 1.15 (A Variant). Suppose that upon meeting one of the two brothers of the above problem, you are not interested in knowing whether he is Arthur or Robert, but rather whether it is Arthur or Robert who is the truthful one. What three-word yes/no question can determine this?

I would ask “Are you Arthur?” If Arthur lies, I will be told “no,” regardless of whether I am speaking to Arthur or Robert. Similarly, if Arthur tells the truth, I will be told “yes” in both cases.

Problem 1.16 (A More Ambitious Task). What about a single yes/no question that would determine both whether he is Arthur or Robert and also whether he is truthful or lies?

There is no such question. The reason is that you are trying to determine which of the four possible cases you are in with only a single binary piece of information. In other words, no matter the question you ask, you will find yourself with multiple cases that return an answer of “yes” or “no.”

Problem 1.17. One day I visited the Island of Knights and Knaves and met an unknown inhabitant who made a statement. I thought for a moment and then said: “You know, before you said that, I had no way of knowing whether it was true or false, but now that you have said it, I know that it must be false, and hence you must be a knave!” What did the inhabitant say?

They said “Everybody on this island is a knave.” This could have been true before it was said, but if a knight could never say that sentence, because it would imply they are a knave. Thus the inhabitant must have been a knave.

After thinking about it, I realized that this doesn’t work if there is only one person on the island; this is not a statement that could have been uttered by a knight or a knave. What is differentiates “Everybody on this island is a knave from “I am a knave” is that the speaker is saying two things, namely that they are a “knave” and that everybody else is too. So we can replace our statement with “I am a knave and I like sports.” This could never have been said by a knight, so we know that the inhabitant must be a knave.
Problem 1.18 (Another Special One). On another occasion during my visit, I met an inhabitant named Al who said: “My father once said that he and I are of different types – that one of us is a knight and the other a knave.”

Is it possible that the father really said that? (The father is also an inhabitant of the island.)

Suppose the father said that, then Al is telling the truth, in other words Al is a knight. If the father is a knight, then it is true that he is of a different type than his son, so he is a knave, a contradiction. But if the father is a knave, then his claim was a lie, i.e. Al and his father are of the same type, but this implies Al’s father is a knight, another contradiction! Therefore the father cannot have said that, so Al is a knave.

Problem 1.19 (Enter a Spy!). One day a spy secretly entered the island. Now, the spy was neither a knight nor a knave; he could sometimes lie and sometimes tell the truth and would always do whatever suited most his convenience. It was known that a spy was on the island, but their identity was not known. After a fruitless hunt, the island police (who are all knights) gave up, hence Inspector Craig of Scotland Yard had to be called in to catch the spy. Well, Craig finally found out that the spy was living with two friends, one of whom was a knight and the other a knave. The three were arrested and brought to trial. It was not known which of the three was the knight, which was a knave, and which was the spy. But Craig realized that by asking only two yes/no questions (not necessarily to the same one of the three) he could finger the spy.

What two questions would work?

Call the three suspects A, B, and C. I would begin by asking A “If I asked you if B is a spy, what would you say?” Suppose the answer is “yes.” There are two cases to consider. If B is actually the spy, then C is either a knight or a knave. If B is not the spy, then A must be the spy, because a knight or knave would not have answered “yes” to that question. Hence, we again have that C is a knight or a knave.

Suppose the answer is “no.” Then we know that B is not the spy, because a knight or a knave would have said “yes.”

Regardless of whether we were told “yes” or “no” by A, we have identified one of the three suspects as either a knight or a knave. Suppose we know that B is a knight or a knave (i.e. the answer was “no”). Then I would ask B “If I asked you whether if A is a spy, what would you say?” If the answer is “yes,” then I know that A is the spy, and if the answer is “no,” I know that C is the spy.

If I knew that C was either a knight or a knave, I would do the same thing as above.

Problem 1.20. One day Abercrombie came across two brothers named Andrew and Bernard. Andrew said: “Both of us are knights.” Abercrombie then asked Bernard: “Is that really true?” Bernard Answered him (he either said yes or no), and Abercrombie then knew what type each was.

At this point, you have enough information to know the type of each. What type is each?

If Bernard had said “yes,” then Abercrombie wouldn’t have had enough information to know whether they were both knights or both knaves (either is case is logically consistent). Since Abercrombie was able to figure out what type each was, we know that the answer must have been “no.” Suppose Andrew is a knight, then so is Bernard. But Bernard claims that they are not both knights, a contradiction. Therefore Andrew is a knave and Bernard is a knight.

Problem 1.21 (Which is Witch?). The island also has a witch doctor, and Abercrombie was curious to meet him. He finally found out that the witch doctor was one of two brothers named Jal and Tak. First Jal said: “I am a knight and my Brother Tak is a knave.” Abercrombie then asked: “Are you the witch doctor?” Jal replied yes. Then Abercrombie asked Tak: “Are you the witch doctor?” Tak answered (either yes or no) and Abercrombie then knew which one was the witch doctor. Which one was it?

If the answer is “yes,” then Abercrombie does not have enough information to solve the problem. Indeed, if Tak is a knight, then he is the witch doctor, and Jal is a knave. Both of his statements were lies, so this is logically consistent. If Tak is a knave, then he is not the witch doctor, which makes Jal the witch doctor and a knight. Jal’s statements were both truths, so this scenario is consistent as well.

We then know that the answer must have been “no.” If Tak is a knave, then he is the witch doctor, which makes Jal a knave. This is all consistent so far. If Tak is a knight, then he is not the witch doctor, making Jal a knight. But Jal claims that Tak is a knave, a contradiction. Hence, Tak must be the witch doctor.
Problem 1.22 (Innocent or Guilty?). Before leaving the island, Abercrombie attended the trial of a native named Snark who was suspected of having committed a robbery. In court were two witnesses named Ark and Bark. The judge (who was truthful) first asked Ark: “Is Snark innocent?” Ark replied: “He once claimed that he is innocent.” Then Bark said: “He once claimed that he is guilty,” at which point Snark said: “Bark is a liar!” The judge then asked Snark: “And what about Ark? Is he a liar?” Snark answered, either yes or no, and the judge (who was good at logic) then knew whether Snark was innocent or guilty.

Which was he?
Suppose that Snark answered “yes.” If Snark is a knight, then Ark and Bark are both liars, which means that Snark never claimed to be innocent or guilty. There is not enough information here to determine whether Snark is guilty or not.
Suppose Snark answered “no.” If Snark is a knight, then Ark is a knight, and Bark is a liar. In this case, Ark’s claim that Snark once claimed to be innocent is true, and since Snark is a knight, Snark is innocent. If Snark is a knave, then Bark is a knight, which means that Snark truly once claimed to be guilty. But since Snark is a knave, he must have been innocent.

2. Male or Female?
Problem 2.1 (The First Test). Abercrombie was led into an auditorium, and an inhabitant of the island came on the stage wearing a mask. Abercrombie was to determine whether the inhabitant was male or female. He was allowed to ask only one yes/no question, and the inhabitant would then write his or her answer on the blackboard (so as not to be betrayed by his or her voice). At first, Abercrombie thought of using a modification of the Nelson Goodman principle, but suddenly realized that he could ask a much simper and less convoluted question that would work. He asked a direct yes/no question that involved only four words. What question would work?
I don’t know if this is what he had in mind, but “Do your men lie?” works. Really, I tried to pack the question “Does a man of your type lie?” into four words. If I am speaking to a male, I will be told “no.” If I am speaking to a woman, I will be told “yes.”
After checking the solution, I realized that his answer is simply “Are you a knight?” Indeed, men will always say “yes,” and women will always say “no.”

Problem 2.2. The inhabitant then left the stage and a new masked inhabitant appeared. This time Abercrombie’s task was to find out, not the sex of the inhabitant, but whether the inhabitant was a knight or a knave. Abercrombie then asked a three-word yes/no question that accomplished this. What question would do this?
“Are you male?” works, since a knight will always answer “yes,” while a knave will always say “no.”

Problem 2.3. Next, a new inhabitant appeared on the stage, and Abercrombie was to determine whether or not the inhabitant was married, but he could ask only one yes/no question. What question should he ask? (The solution involves but a minor modification of the Nelson Goodman principle. Indeed, Abercrombie could find out any information he wanted by asking just one yes/no question.)
“If I asked you ‘Are you married?’ what would you say?” would work. Indeed, if the inhabitant is truthful, they will answer truthfully (obviously). If the inhabitant only tells lies, then they would lie when asked “Are you married?” But then they would answer the opposite to the question “If I asked you ‘Are you married?’ what would you say?”

Problem 2.4. In the next test, an inhabitant wrote a sentence from which Abercrombie could deduce that the inhabitant must be a female knight. What sentence could do this?
“I am a male knave.” Neither a male knight nor a female knave would write this sentence because they are truthful. A male knave also would not write this, because then he would be telling the truth. So the inhabitant must have been a female knight.

Problem 2.5. Next, a sentence was written from which Abercrombie could deduce that the writer must be a male knight. What could this sentence have been?
The sentence could have been “I am male or I am a female knight.” Neither a male knave nor a female
knight could never say this because it would be true. A female knave would not say this because then it would be a lie. Thus, the inhabitant must have been a male knight.

**Problem 2.6.** Then an inhabitant wrote a sentence from which it could be deduced that the writer must be either female or a knight (or maybe both), but there was no way to tell which. What sentence would work?

The sentence “I am male or I am a knave” works. The writer could be a female knave or a male knight, so there is no way to determine whether the inhabitant is female or a knight. Moreover, if the writer is not a knight, then they must be female.

**Problem 2.7.** When Abercrombie returned from the island, he told his friend, a lawyer, that the day before he left, he had witnessed an extremely unusual trial! A valuable diamond had been stolen, and it was not known whether the thief was male or female, or whether a knight or a knave. Three suspects were tried in court, and it was known that one of the three was the thief, but it was not known which one. The strangest thing about the trial was that all three appeared masked, since they did not want their sex to be known. [They had the legal right to do this!] The also refused to speak, since they did not want their voices to betray their sex, but they were willing to write down answers to any questions asked of them. The judge was a truthful person. We will call the three defendants A, B, and C, and here are the questions asked by the judge and the answers they wrote:

```
Judge (to A):  What do you know about all this?
    A:  The thief is male.
Judge:  Is he a knight or a knave?
    A:  He is a knave.
Judge (to B):  What do you know about A?
    B:  A is female.
Judge:  Is she a knight or a knave?
    B:  She is a knave.
```

The judge thought for a while and then asked C: “Are you, by any chance, the thief?” C then wrote down his or her answer, and the judge then knew which one was the thief.

At this point of Abercrombie’s account of the trial, his friend asked: “What answer did C write?”

“Unfortunately, I don’t remember,” replied Abercrombie. “C either wrote yes or wrote no, but I don’t remember which. At any rate, at this point the judge knew which one was the thief.”

Well, at this point, you, the reader, have enough information to solve the case! Which of the three was the thief?

I think that the question is poorly worded. Answers like “He is a knave” don’t really make sense if the thief was female, in my opinion. At best, I think they should be classified as a lie. I think a better answer would be “The thief is a knave.”

In any case, suppose that the answer was “yes.” If C is truthful, then A must be a liar, which means that B is a liar as well. But then A is a male knight, which means he is truthful, a contradiction. On the other hand, if C is lying, B must be the thief for the following reason: if A is the thief, then A cannot be telling the truth, because then he would be claiming to be a male knave. Then A must be a female knight. But then one B’s answers is a truth while the other is a lie, a contradiction. Hence B is the thief. Indeed, this is possible if A is a male knight, B is a male knave, and C is a female knight.

If C had said no, there would be no way of solving the case.

### 3. Silent Knights and Knaves

**Problem 3.1.** Abercrombie, who knew the rules of this island, decided to pay a visit. He met a native and asked him: “Does a red card signify yes?” The native then showed him a red card.

Form this, is it possible to deduce what a red card signifies? Is it possible to deduce whether the native was a knight or a knave?

A knight will always raise a red card to this question, while a knave will always raise a black card. Therefore, there is no way of determining what the cards signify, but Abercrombie now knows that the native is
a knight.

**Problem 3.2.** Suppose one wishes to find out whether it is a red card or a black card that signifies yes. What simple yes/no question should one ask?

Any native will answer “yes” to the question “Are you a knight?” So we can use this question to determine the meaning of the cards.

**Problem 3.3.** Suppose, instead, one wishes to find out whether the native is a knight or a knave. What yes/no question should one ask?

As in problem 3.1, we can ask “Does a red card signify yes?” A knight will raise a red card, while a knave will raise a black one.

**Problem 3.4.** Without knowing what red and black signify, what yes/no question is such that the native would be sure to flash a red card?

The question “How would you answer the question ‘Does a red card signify yes?’?” works. If the native is a knight, we already know that they will always raise a red card here. Suppose the islander is a knave. First, suppose the red card signifies yes. If we asked “Does a red card signify yes?” they would answer “no,” and lift a black card. Therefore, to our original question, they would answer “yes,” and raise the red card. Now suppose that the red card signifies “no.” If we asked “Does the red card signify yes?” they would say “yes.” Therefore, the knave will answer “no” to our original question, and raise a red card.

**Problem 3.5.** While Abercrombie was on the island, he attended a curious trial: A valuable diamond had been stolen. A suspect was tried, and three witnesses $A$, $B$, and $C$ were questioned. The presiding judge was from another land and didn’t know what the colors red and black signified. Since non-inhabitants were present at the trial, the three witnesses were willing to answer questions only in their sign language of red and black.

First, the judge asked $A$ whether the defendant was innocent. $A$ responded by flashing a red card.

Then, the judge asked the same question of $B$, who then flashed a black card.

Then, the judge asked $B$ a second question: “Are $A$ and $C$ of the same type?” (meaning both knights or both knaves). $B$ flashed a red card.

Finally, the judge asked $C$ a curious question: “Will you flash a red card in answer to this question?” $C$ then flashed a red card.

Is the defendant innocent or guilty?

Either red means yes and $C$ is a knight, or red means no and $C$ is a knave. We also know that $A$ and $B$ must be of different types, since they answered differently to the same question.

- Case 1. Red means yes and $C$ is a knight.
  - Case 1.1. $A$ is a knight. Then $B$ is a knave. Moreover, $C$ is a knave, a contradiction.
  - Case 1.2. $A$ is a knave. Then $B$ is a knight, which means that $C$ is also a knave, another contradiction.

- Case 2. Red means no and $C$ is a knave.
  - Case 2.1. $A$ is a knight. Then $B$ is a knave. Then $C$ is a knight, a contradiction.
  - Case 2.2. $A$ is a knave. Then $B$ is a knight, which means that $C$ is a knight, a contradiction.

Every single case leads to a contradiction, so I think that there is a mistake in the question. In his proof in the solutions, he proves that the case where the defendant is guilty cannot occur, but he does not prove that the case where the defendant is innocent can occur.

4. Mad or Sane?

**Problem 4.1 (Let’s Be Careful!).** We shall start with a very tricky puzzle, but also one that illustrates a basic principle. Is it possible for an inhabitant of this island to say: “I believe I am mad”?
A sane person could not say this, because then they would incorrectly believe they are mad. However, a mad person could say this. Their belief that they believe they are mad is can be incorrect. This is different from them saying “I am mad.”

**Problem 4.2 (A Simple Form of the Nelson Goodman Principle).** The Nelson Goodman principles for the islands 1 and 2 involve questions that are rather convoluted and unnatural. Well, on the present island, there is a much more natural-sounding yes/no question you could ask to ask to ascertain any information you want. For example, if you want to find out whether a given native is married or not, there is a relatively natural-sounding question you can ask— one, in fact, having only six words. What question would work?

The question “Do you believe you are married?” works. Indeed, if you are speaking to a sane person, they will answer truthfully. If they are insane and married, then they won’t believe they are married, but they will believe that they believe they are married. Therefore, the native, regardless of their sanity, will answer in accordance to reality.

**Problem 4.3.** When Abercrombie got settled on this island, he first interviewed three siblings named Henry, Dianne, and Maxwell. Henry and Dianne made the following statements:

- **Henry:** Maxwell believes that at least one of us is mad.
- **Dianne:** Maxwell is sane.

What type is each?

Suppose that Henry is sane. Then Maxwell must be sane as well, otherwise he would be mad and correctly believe that at least one of the siblings is mad. But then Dianne is sane as well. This is a contradiction, since Maxwell incorrectly believes there is at least one mad sibling, meaning he is mad.

Then Henry must be mad. Then Maxwell does not believe any of the siblings are mad, which is incorrect. Then Maxwell is mad. Then Dianne is mad. They are all mad!

**Problem 4.4.** Next, Abercrombie interviewed Mary and Gerald, a married couple, together with their only child, Lenore. Here is the dialogue that took place.

- **Abercrombie (to Gerald):** I heard that your wife once said that all three of you are mad. Is that true?
  - **Gerald:** No, my wife never said that.
- **Abercrombie (to Lenore):** Did your father once say that exactly one of you three is sane?
  - **Lenore:** Yes, he once said that.
- **Abercrombie (to Mary):** Is your husband sane?
  - **Mary:** Yes.

What type is each?

Suppose Mary is mad, then so is Gerald. That means that Mary must have said that all three of them were mad. Hence, Lenore must be sane, which means that Gerald actually said exactly one of them is sane. But this last statement is true, contradicting the assumption that Gerald is mad. Hence Mary must be sane.

If Mary is sane, then so is Gerald. Then Lenore must be mad, since Gerald would not say that exactly one of the three is sane given that Gerald and Mary are both sane in this case. In conclusion, Mary and Gerald are sane, while Lenore is mad.

**Problem 4.5.** Abercrombie’s next interview was a bit more puzzling. He met a married couple, Arthur and Lillian Smith. Arthur was the only one who said anything, and what he said was: “My wife once said that I believe that she believes I am mad.”

What can be deduced about either one?

Say Arthur and Lillian are both sane. Then Arthur believes that Lillian believes he is mad. But this belief is incorrect, since Lillian does not believe Arthur is mad as he is sane. This is a contradiction, so Arthur and Lillian cannot both be sane.

Say Arthur is sane and Lillian is mad. Then Lillian said what Arthur claims she said. Then, since Lillian is mad, Arthur does not believe that Lillian believes he is mad. This belief (or disbelief) is incorrect, since
the mad Lillian incorrectly believes the sane Arthur is mad. This contradicts the assumption that Arthur is sane.

Then Arthur must be mad. Nothing can be deduced about Lillian, because we have no information about her (since mad Arthur’s statement is incorrect).

**Problem 4.6.** Abercrombie next interviewed eight brothers named Arthur, Bernard, Charles, David, Ernest, Frank, Harold, and Peter. They made the following statements:

- **Charles:** Arthur is mad.
- **David:** Bernard and Charles are not both mad.
- **Ernest:** Arthur and David are alike, as far as sanity goes.
- **Frank:** Arthur and Bernard are both sane.
- **Harold:** Ernest is mad.
- **Peter:** Frank and Harold are not both mad.

*From this confusing tangle, it is possible to determine the madness or sanity of one of the eight. Which one, and what is he?*

Suppose that Frank is sane. Then so are Arthur and Bernard. That makes Charles mad and David sane. Ernest is then sane as well, making Harold mad. Peter is also sane. We have the following list:

Arthur : sane
Bernard : sane
Charles : mad
David : sane
Ernest : sane
Frank : sane
Harold : mad
Peter : sane

Suppose that Frank is mad. Then either Arthur or Bernard is mad (or both). Suppose Arthur is mad. Then Charles is sane, and so is David. Then Ernest is mad, making Harold sane. Peter is then sane. Bernard could be either sane or mad. We have the following list:

Arthur : mad
Bernard : ?
Charles : sane
David : sane
Ernest : mad
Frank : mad
Harold : sane
Peter : sane

Now suppose Arthur is sane. Then Bernard must be mad. Charles is also mad, making David mad, and Ernest as well. Harold is then sane, and so is Peter. We have the following list:

Arthur : sane
Bernard : mad
Charles : mad
David : mad
Ernest : mad
Frank : mad
Harold : sane
Peter : sane

In our three lists, the only person who was always the same type is Peter, so we know that Peter must be sane.
Problem 4.7 (A Metapuzzle) Before Abercrombie left the island, one of the sane inhabitants, whose name was David, told him of a trial he had attended some time ago. The defendant was suspected of having stolen a watch. First, the judge (who was sane) said to the defendant: “I have heard that you once claimed that you stole the watch. Is that true?” The defendant answered (either yes or no). Then the judge asked: “Did you steal the watch?” The defendant then answered (either yes or no) and the judge then knew whether he was innocent or guilty.

“What answers did the defendant give?” asked Abercrombie.

“I don’t quite remember,” replied David. “It was quite some time ago. I do, however, remember that he didn’t answer no both times.”

Was the defendant innocent or guilty?

Suppose the answers were “yes” and “yes.” Suppose that the defendant is sane. Then the defendant stole the watch. Suppose that the defendant is mad. Then they did not steal the watch. There are no contradictions in either case, so there is no way of determining if the defendant is innocent or guilty. Thus, the answers were not “yes” and “yes.”

Suppose the answers were “no” and “yes.” If the defendant is sane, then they never said they stole the watch, but they actually did. If the defendant is mad, then they did not steal the watch, but they did once say that they did, which is consistent. Again, there is no way of determining if the watch was stolen.

Suppose the answers were “yes” and “no.” Then defendant claims to not have stolen the watch, but to have said they did. The defendant must be mad, which means they did steal the watch.

5. The Difficulties Double!

Problem 5.1. Suppose you meet a native of this island and want to know whether he is sane or mad. What single yes/no question could determine this?

The inhabitants that make true statements are sane knights and mad knaves. We will say that these inhabitants are of type 1. The inhabitants that make false statements are mad knights and sane knaves. We will say that these inhabitants are of type 2. Thus, a question that works is “Are you of type 1 if and only if you are sane?” Suppose the answer is “yes.” Then then the inhabitant is either of type 1 and sane, or of type 2 and sane. In either scenario, they are sane. If the answer is “no,” similar reasoning shows that they must be mad.

Problem 5.2. Suppose, instead, you wanted to find out whether he was a knight or a knave.

For the same reason as in problem 5.1, the question “Are you of type 1 if and only if you are a knight?” works.

Problem 5.3. What yes/no question could you ask that would ensure they will answer yes?

“Are you of type 1?”

Problem 5.4. There is a Nelson Goodman principle for this island; one can find out any information one wants with just one yes/no question. For example, suppose you wanted to know whether there is gold on this island. What single yes/no question would you ask?

I was lazy and used the Nelson Goodman principle in problems 5.1 and 5.2. The question to ask is “Are you of type 1 if and only if there is gold on this island?”

Problem 5.5. When Abercrombie arrived on this island, he met a native named Hal who made a statement from which Abercrombie could deduce that he must be a sane knight. What statement would work?

He might have said “I am sane or I am a mad knight.” The statement could not have been said by a mad knight or a sane knave, because then a type 2 inhabitant would be telling the truth. Moreover, the statement could not have been said by a mad knave because the a type 1 inhabitant would be telling a lie. Therefore, Hal must be a sane knight.

Problem 5.6. Abercrombie and Hal became fast friends. They often went on walks together, and Hal was sometimes quite useful in helping Abercrombie in his investigations. On one occasion, they spied two inhabitants walking toward them.
"I know them!" said Hal. "They are Auk and Bog. I know that one of them is sane and the other is mad, but I don't remember which one is which."

When the two came up to them, Abercrombie asked them to tell him something about themselves. Here is what they said:

Auk: Both of us are knaves.
Bog: That is not true!

Which of the two is mad?
Suppose that Auk is sane and Bog is mad. Auk cannot be a knight, because then he would truthfully be claiming to be a knave. So Auk is a sane knave. Thus, Bog must be a mad knight. But this cannot be, because his statement is true.

Then it must be that Auk is mad and Bog is sane. Indeed, if Auk is a mad knight and Bog is a sane knight, their statements are consistent with their type.

Problem 5.7. On another occasion, they came across two other natives named Ceg and Fek. Hal told Abercrombie that he remembered that one was sane and one was mad, but wasn't sure which was which. He also remembered that one was a knight and the other a knave, but again was not sure which was which.

"As a matter of fact," said Abercrombie, "I came across these same two natives a couple of days ago and Ceg said that Fek is mad and Fek said that Ceg was a knave."

"Ah, that settles it!" said Hal.
Hal was right. What type is each?
Suppose that Ceg is sane and Fek is mad. According to Abercrombie (who I will assume is truthful), Ceg made a true statement, and must therefore be a sane knight. Fek, on the other hand, made a false statement, and must therefore be a mad knight. This cannot be, because one of them is a knight and the other is a knave.

Suppose that Ceg is mad and Fek is sane. Ceg then made a false statement, which means he must be a mad knight. Fek, also made a false statement, making him a sane knave.

Problem 5.8. On another occasion, Abercrombie and Hal came across two natives named Bek and Drog, who made the following statements:

Bek: Drog is mad.
Drog: Bek is sane.
Bek: Drog is a knight.
Drog: Bek is a knave.

What type is each?
Suppose Drog is mad. Then Bek makes true statements, which means that Drog is a mad knight, meaning he makes false statements. Thus, Bek is a mad knight. But Bek makes true statements, so this cannot be.

Suppose Drog is sane. Then Bek makes false statements, which means that Drog is a sane knave. Therefore Bek is a mad knight.

Problem 5.9 (A Metapuzzle). Several days later, after Abercrombie learned more about some of the natives, he and Hal were walking along one late afternoon and spied a native mumbling something to himself.

"I know something about him," said Abercrombie. "I know whether he is a knight or a knave, but I don't know whether he is sane or mad."

"That's interesting!" said Hal. "I, on the other hand, happen to know whether he is sane or mad, but I don't know whether he is a knight or a knave."

When the two got closer, they heard what the native was mumbling, which was "I am not a sane knave." The two both thought for a while, but Abercrombie still didn't have enough information to determine whether the native was sane or mad, nor did Hal have enough information to determine whether the native was a knight or a knave.

At this point, you have enough information to determine the type of the native. Was he sane or mad, and was he a knight or a knave?
Suppose Abercrombie knew he was a knight. Then he would know that he is a sane knight, because a mad knights make false statements. Thus the native is a knave.

Suppose Hal knew he was mad. Then he would know that he is a mad knight, because a mad knave would not correctly claim to not be a sane knight. Hence, the native is sane.

Thus, the native is a sane knave.

6. A Unification

**Problem 6.1.** Is there a Nelson Goodman-type principle for this crazy island? That is, can one find out any information one wants by asking just one yes/no question? For example, suppose you visit the island and want to know if there is gold on it. You meet a masked native and don’t know the sex of the native, nor whether he or she is a knight or a knave, nor whether made or sane, nor what the colors red and black signify to him or her. Is there a single yes/no question that you could ask to determine whether there is gold on the island, or is no such question possible?

We will say that a speaker is truthful when they answer in accordance to reality.

The following convoluted question works: “Do you lift a red card when the true answer is ‘yes’ if and only if there is gold on this island?” To see that this question works, we need to check a few cases.

- **Case 1:** There is gold on the island.
  - **Case 1.1:** The native is truthful.
    - *Case 1.1.1:* The red card means yes. Then the native will lift a red card.
    - *Case 1.1.2:* The red card means no. The native does not lift a red card to signify yes, so they will answer no to the question, which means they will lift a red card.
  - **Case 1.2:** The native is not truthful.
    - *Case 1.2.1:* The red card means yes. The native does not lift a red card when the true answer is yes, so the true answer to the question is no. Since the native is not truthful, they will answer yes, i.e. raise a red card.
    - *Case 1.2.2:* The red card means no. In this case, the native lifts a red card when the true answer is yes, so the true answer to the question is yes. The untruthful native will then answer no, and raise a red card.

- **Case 2:** There is no gold on the island. We repeat all the sub-cases treated above. All the outcomes will get flipped due to the fact that there is no gold on the island, so no matter the type of the native and the meaning of the cards, the native will answer by raising a black card.

Thus, if the native answers by raising a red card, we know that there is gold on the island. If a black card is raised, there is no gold on the island.

**Problem 6.2.** A related problem is this: Is there a yes/no question that will ensure that the native addressed will respond by flashing a red card?

The following question works: “Do you lift a red card when the true answer is ‘yes’?”

**Problem 6.3.** We have considered the Nelson Goodman principle in six different situations: the island of knights and knaves, the knight/knave island in which males and females respond differently, the island of knights and knaves where natives respond by flashing red or black cards instead of saying yes or no, the island of the sane and the mad, the island that combines knight and knaves with sane and mad, and, finally, the crazy island of this chapter. We have given different Nelson principles for four of them and stated that there is also one for the island of Chapter 3. Is it not possible to unify them into one principle that is simultaneously applicable to all six cases, and possibly others?

Yes, there is! Can the reader find one? (There is not just one that works; different readers may well find different general principles, all of which fit the bill. We provide one in the solutions.

Suppose you are on an island where there are two responses to yes/no questions: X and Y. And suppose that some islanders respond by X to mean yes and by Y to mean no, and other islanders respond by Y to
mean yes and by $X$ to mean no, but you do not know which are which. Then, to ascertain whether any statement is true, you could ask any native “do you respond by $X$ when the true answer is ‘yes’ if and only if [insert statement] is true?” If they respond by $X$, you know the statement is true, and if they respond by $Y$, you know the statement is false.

**Part II – Be Wise, Symbolize!**

**7. Beginning Propositional Logic**

**Exercise 7.1.** State which of the following are tautologies, which are contradictions and which are contingent (sometimes true, sometimes false).

(a) $(p \implies q) \implies (q \implies p)$. Contingency.

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

(b) $(p \implies q) \implies (\neg p \implies \neg q)$. Contingency.

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

(c) $(p \implies q) \implies (\neg q \implies \neg p)$. Tautology.

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

(d) $(p \equiv q) \equiv (\neg p \equiv \neg q)$. Tautology.

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

(e) $p \implies \neg p$. Contingency.

<table>
<thead>
<tr>
<th>$p$</th>
<th>$\neg p$</th>
<th>$p \implies \neg p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T$</td>
<td>$F$</td>
<td>$F$</td>
</tr>
<tr>
<td>$F$</td>
<td>$T$</td>
<td>$T$</td>
</tr>
</tbody>
</table>

(f) $p \equiv \neg p$. Contradiction.

<table>
<thead>
<tr>
<th>$p$</th>
<th>$\neg p$</th>
<th>$p \equiv \neg p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$T$</td>
<td>$F$</td>
<td>$F$</td>
</tr>
<tr>
<td>$F$</td>
<td>$T$</td>
<td>$F$</td>
</tr>
</tbody>
</table>
(g) \( \sim (p \land q) \equiv (\sim p \land \sim q) \). Contingency.

\[
\begin{array}{c|c|c|c|c|c|c}
  & p & q & \sim p & \sim q & p \land q & \sim (p \land q) \\
  \hline
  T & T & T & F & F & T & F \\
  T & F & F & T & F & T & F \\
  F & T & T & F & F & T & F \\
  F & F & T & T & T & T & T \\
\end{array}
\]

(h) \( \sim (p \land q) \equiv (\sim p \lor \sim q) \). Tautology.

\[
\begin{array}{c|c|c|c|c|c|c}
  & p & q & \sim p & \sim q & p \lor q & \sim (p \lor q) \\
  \hline
  T & T & T & F & F & T & F \\
  T & F & F & T & F & T & F \\
  F & T & T & F & F & T & F \\
  F & F & T & T & T & T & T \\
\end{array}
\]

(i) \( (\sim p \lor \sim q) \Rightarrow \sim (p \lor q) \). Contingency.

\[
\begin{array}{c|c|c|c|c|c|c}
  & p & q & \sim p & \sim q & p \lor q & \sim (p \lor q) \\
  \hline
  T & T & F & F & F & T & T \\
  T & F & F & T & F & T & F \\
  F & T & T & F & F & T & F \\
  F & F & T & T & T & T & T \\
\end{array}
\]

(j) \( \sim (p \lor q) \equiv (\sim p \land \sim q) \). Tautology.

\[
\begin{array}{c|c|c|c|c|c|c}
  & p & q & \sim p & \sim q & p \lor q & \sim (p \lor q) \\
  \hline
  T & T & T & T & F & F & T \\
  T & F & F & T & F & F & T \\
  F & T & T & T & F & F & T \\
  F & F & T & T & T & T & T \\
\end{array}
\]

(k) \( (\sim p \lor \sim q) \land (p \equiv (p \Rightarrow q)) \). Contradiction.

\[
\begin{array}{c|c|c|c|c|c|c|c}
  & p & q & \sim p & \sim q & \sim p \lor \sim q & p \Rightarrow q & p \equiv (p \Rightarrow q) \\
  \hline
  T & T & F & F & F & T & T & F \\
  T & F & F & T & F & T & F & F \\
  F & T & T & F & T & F & F & F \\
  F & F & T & T & T & F & F & F \\
\end{array}
\]

(l) \( (p \equiv (p \land q)) \equiv (q \equiv (p \lor q)) \). Tautology.

\[
\begin{array}{c|c|c|c|c|c|c|c}
  & p & q & p \land q & p \lor q & p \equiv (p \land q) & q \equiv (p \lor q) \\
  \hline
  T & T & T & T & T & T & T \\
  T & F & F & T & F & F & T \\
  F & T & F & T & T & T & T \\
  F & F & F & F & T & T & T \\
\end{array}
\]

**Problem 7.1.** What is the method? (See the text for context.)

Let \( p_1, \ldots, p_n \) be a list of propositions, and suppose that the truth table for the \( p_i \) has been filled in (this will be an \( n \times 2^n \) table). Append an extra column to the right of the truth table, and fill it with an arbitrary sequence of \( T \)'s and \( F \)'s. Our goal is to construct a formula \( X \) involving the \( p_i \) such that the extra column of the truth table corresponds to the truth values of \( X \).

For each \( T \) in the extra column, for the formula \( p_1^i \land \cdots \land p_n^i \), where \( p_i^i \) is \( p_i \) if the corresponding truth value is \( T \) in the truth table, and \( \sim p_i \) if it is \( F \). Thus we get a finite sequence \( X_1, \ldots, X_m \) of formulas involving the \( p_i \). Now just let \( X = X_1 \lor \cdots \lor X_m \).

In the example given in the text, the formula would be \( X = (p \lor \sim q \land r) \lor (\sim p \land q \land \sim r) \lor (\sim p \land \sim q \land \sim r) \).
8. Liars, Truth-Tellers, and Propositional Logic

Problem 8.1 (An If-Then Problem). Suppose you visit the island of knights and knaves and wish to find out some information such as whether there is gold on the island. You meet a native and ask him whether there is gold on the island. All he replies is: "If I am a knight, then there is gold on the island." From this, is it possible to tell whether there is gold, and whether he is a knight or a knave? Yes, it is possible to determine both! How? (The solution is important!)

Let \( k \) be the proposition that the native is a knight and let \( g \) be the proposition that there is gold on the island. Then we know that \( k \equiv (k \Rightarrow g) \). The truth table for this formula is

<table>
<thead>
<tr>
<th>( k )</th>
<th>( g )</th>
<th>( k \Rightarrow g )</th>
<th>( k \equiv (k \Rightarrow g) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>T</td>
<td>F</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>T</td>
<td>F</td>
</tr>
</tbody>
</table>

So the only scenario is that the native is a knight, and there is gold on the island.

Problem 8.2 (A Variant). Suppose that the native had instead said: "Either I am a knave or there is gold on the island." Is it now possible to determine whether there is gold there, or the type of the native?

We construct the truth table for \( k \equiv (\sim k \lor g) \):

<table>
<thead>
<tr>
<th>( k )</th>
<th>( g )</th>
<th>( \sim k )</th>
<th>( \sim k \lor g )</th>
<th>( k \equiv (\sim k \lor g) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>F</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>T</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>F</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>T</td>
<td>T</td>
<td>F</td>
</tr>
</tbody>
</table>

Hence, the native is a knight and there is gold on the island.

Problem 8.3 (Another Variant). Suppose that the native had instead said: "I am a knave and there is no gold on the island." What can be deduced.

We construct the truth table for \( k \equiv (\sim k \land \sim g) \):

<table>
<thead>
<tr>
<th>( k )</th>
<th>( g )</th>
<th>( \sim k )</th>
<th>( \sim g )</th>
<th>( \sim k \land \sim g )</th>
<th>( k \equiv (\sim k \land \sim g) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>F</td>
<td>T</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>T</td>
<td>F</td>
<td>F</td>
<td>T</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>F</td>
</tr>
</tbody>
</table>

Hence, the native is a knave and there is gold on the island.

Problem 8.4 (Gold or Silver?). Suppose that a native makes the following statement: "There is gold on this island, and if there is gold, then there is also silver." Can it be deduced whether he is a knight or a knave. Can it be determined whether there is gold? What about silver?

Let \( s \) be the statement that there is silver on the island. Construct the truth table for \( k \equiv (g \land (g \Rightarrow s)) \):

<table>
<thead>
<tr>
<th>( k )</th>
<th>( g )</th>
<th>( s )</th>
<th>( g \Rightarrow s )</th>
<th>( g \land (g \Rightarrow s) )</th>
<th>( k \equiv (g \land (g \Rightarrow s)) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>T</td>
<td>T</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>T</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>T</td>
<td>F</td>
<td>F</td>
<td>T</td>
</tr>
<tr>
<td>T</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>T</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>F</td>
<td>F</td>
</tr>
<tr>
<td>F</td>
<td>T</td>
<td>F</td>
<td>F</td>
<td>F</td>
<td>T</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>T</td>
<td>F</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>F</td>
<td>F</td>
<td>T</td>
<td>T</td>
<td>T</td>
<td>T</td>
</tr>
</tbody>
</table>
Problem 8.5 (Gold or Silver?). Suppose that he instead makes the following two separate statements:

1. There is gold on this island.
2. If there is gold here, then there also is silver.

Is the solution the same as that of the last problem?

Either (1) and (2) are both true, or they are both false. Hence, we know that \( g \equiv (g \Rightarrow s) \) is true. If \( g \) is false, \( g \Rightarrow s \) is necessarily true. Therefore, \( g \) must be true. Then \( g \Rightarrow s \) is also true, so \( s \) must be true. Moreover, \( k \equiv g \), so \( k \) is true as well.

Hence, the native is a knight and there is gold and silver on the island.

Problem 8.6 (A Metapuzzle). One day a man was tried for a crime on this island. First the prosecutor pointed to the defendant and said: “If he is guilty, then he had an accomplice.” Then the defense attorney said: “That’s not true!” Now, the judge didn’t know whether the prosecutor was a knight or a knave, nor did he know the type of the defense attorney. Later, he found out whether the defense attorney was a knight or a knave and was then able to convict or acquit the defendant. Which did he do?

Let \( k_1, k_2, \) and \( k_3 \) be the propositions that the prosecutor, the defendant, and the defense attorney (respectively) are knights. Let \( g \) be the proposition that the defendant is guilty. Let \( a \) be the proposition that the defendant had an accomplice.

If \( \sim k_3 \) is true, then \( g \Rightarrow a \) is true. In this case, \( g \) and \( \sim g \) could be true. However, if \( k_3 \) is true, then we know that \( g \Rightarrow a \) is false. This can only happen when \( g \) is true and \( a \) is false.

Therefore, the judge convicted the defendant.

Problem 8.7. What question could you ask a native such that it is impossible for them to answer either yes or no without violating their type?

A question could be “Are you a knight if and only if your answer to this question will be ‘no’?” Let \( k \) be the proposition that the native is a knight and let \( P \) be the proposition that the answer to the question is yes.

Suppose that the answer to the question is yes. Then \( \sim P \) is true. But then, by the Nelson Goodman principle, \( P \) is also true. Similarly, if the answer is no, then \( P \) is true. But, again by the Nelson Goodman principle, the \( \sim P \) is true.

Problem 8.8. Suppose there is a spy on the island (who is neither a knight nor a knave) and you ask him the question given in the solution to the last problem. Could he answer it truthfully? Could he answer it falsely?

If he answers yes, then that will be a truthful answer, since he did not answer no and he is not a knight. More formally, the answer yes is the assertion that \( k \equiv P \) is true. Indeed, \( k \) and \( P \) are both false in this scenario, so the spy answered truthfully. If he answers no, then the assertion is that \( k \equiv P \) is false. Indeed, \( k \) is false, but \( P \) is true, so \( k \equiv P \) is false. Therefore, the spy can only answer truthfully.

Strange! This is a yes/no question with no correct answer (if you are speaking to someone who is neither a knight nor a knave).

Problem 8.9. In the above trial, suppose \( A_1 \) and \( A_2 \) had made the following statements instead:

\[
A_1 : \quad \text{If } A_2 \text{ is a knight, then the defendant is innocent.} \\
A_2 : \quad \text{If } A_1 \text{ is a knave, then the defendant is innocent.}
\]

What can be deduced?

Let \( k_1 \) and \( k_2 \) be the propositions that \( A_1 \) and \( A_2 \) are knights. Let \( g \) be the proposition that the defendant is guilty. The only way for \( A_1 \) or \( A_2 \) to be knaves is if their ‘if-then’ statements are of the form \( t \Rightarrow f \). Then we know that \( \sim k_1 \Rightarrow k_2 \) and \( \sim k_2 \Rightarrow \sim k_1 \) are true. From the first of these, \( \sim k_2 \Rightarrow k_1 \) is true. The only way for both \( \sim k_2 \Rightarrow \sim k_1 \) and \( \sim k_2 \Rightarrow k_1 \) to be true is if \( k_2 \) is true.
Now, we know that \( \sim k_1 \Rightarrow \sim g \) is true. If \( \sim k_1 \) is true, then so is \( \sim g \). But then \( k_2 \Rightarrow \sim g \) is true, making \( k_1 \) true, contradicting the assumption that \( \sim k_1 \) is true. Thus, \( k_1 \) is true. Moreover, this implies that \( g \) is false.

Thus, the witnesses are knights, and the defendant is innocent.

**Problem 8.10.** Suppose \( A_1 \) and \( A_2 \) had made the following statements instead:

\[
\begin{align*}
A_1 & : \text{If either of us is a knight, then the defendant is guilty.} \\
A_2 & : \text{If either of us is a knave, then the defendant is guilty.}
\end{align*}
\]

What is the solution?

From \( A_1 \)'s statement, we know that \( (\sim k_1 \land \sim k_2) \Rightarrow k_1 \) is true. Therefore, \( \sim k_1 \lor \sim k_2 \) is false, meaning that \( k_1 \lor k_2 \) is true.

Suppose that \( g \) is false. Then \( (k_1 \lor k_2) \Rightarrow g \) is false, implying that \( \sim k_1 \) is true. In turn, this implies that \( \sim k_1 \lor \sim k_2 \Rightarrow g \) is false, which means that \( \sim k_2 \) is true. We conclude that \( \sim k_1 \land \sim k_2 \) is true, contradicting the conclusion we made in the first paragraph that it is false.

Therefore \( g \) must be true. This means that \( k_1 \land k_2 \) are both true. Hence, both witnesses are knights and the defendant is guilty.

**Problem 8.11.** Suppose \( A_1 \) and \( A_2 \) had made the following statements instead:

\[
\begin{align*}
A_1 & : \text{If I am a knight and \( A_2 \) is a knave, then the defendant is guilty.} \\
A_2 & : \text{That is not true!}
\end{align*}
\]

Is the defendant guilty or not?

Suppose \( g \) is false and \( k_1 \) is true. Then \( (k_1 \land \sim k_2) \Rightarrow g \) is true. This means that \( k_1 \land \sim k_2 \) must be false. Then \( k_2 \) is true. But then \( k_2 \Rightarrow \sim k_1 \) is true implies \( k_1 \) is false, a contradiction. Then suppose \( k_1 \) is false. Then \( k_1 \land \sim k_2 \) must be true, a contradiction.

From the above, we conclude that \( g \) must be true. Therefore, \( A_1 \)'s statement is true, so \( k_1 \) is true, so \( k_2 \) is false. In other words, \( A_1 \) is a knight, \( A_2 \) is a knave, and the defendant is guilty.

**Problem 8.12.** Suppose we are again given that \( A_1 \) and \( A_2 \) are a married (heterosexual) couple, and they write the following statements:

\[
\begin{align*}
A_1 & : \text{My spouse is a knight.} \\
A_2 & : \text{We are both knights.}
\end{align*}
\]

From this it is possible to tell both which one is the husband, and which knight-knave type each is. What is the solution? (It can be found either with a little ingenuity or systematically using a truth table.)

Let \( k_1 \) and \( k_2 \) be the propositions that \( A_1 \) and \( A_2 \) (respectively) are knights. Let \( m_1 \) and \( m_2 \) be the propositions that \( A_1 \) and \( A_2 \) (respectively) are male. Suppose that \( k_1 \land k_2 \) is true. Then \( A_1 \) and \( A_2 \) are both male knights (since they are knights and their statements are true). But they are a heterosexual couple, so this cannot be the case. Hence \( k_2 \equiv m_2 \) is false.

Suppose \( k_2 \) is true. Then \( k_1 \) must be false by the above work. But since \( A_1 \) tells the truth in this case, \( m_1 \) must be false. Therefore \( m_2 \) must be true. But then \( m_2 \land k_2 \) is true, which makes \( k_1 \land k_2 \) true. This is a contradiction.

Therefore \( A_1 \) and \( A_2 \) must be liars. So \( k_2 \) is false, implying that \( m_2 \) is true. Hence, \( m_1 \) is false, which makes \( k_1 \) true. In summary, \( A_1 \) is a female knight and \( A_2 \) is a male knave.

**Problem 8.13.** What question could you ask that would make it impossible for the native to flash either red or black?

In the text, a Type 1 inhabitant is defined to be an inhabitant who lifts a red card when the correct answer is yes. A question that could work is “Are you a Type 1 inhabitant if and only if you will answer this question by lifting a black card?”

Indeed, suppose the inhabitant is of type 1. If they raise a black card, then they are saying no to a question whose correct answer is yes. Similarly, if they raise a red card, they are saying yes to a question whose correct answer is no. Both cases lead to contradictions.
Now suppose the inhabitant is of Type 2. Then they lift a black card when the correct answer is yes. If they answer the question by lifting a black card, then they are signifying that the correct answer to the question is yes, but the correct answer is no! Similarly, if they raise a red card, they are saying no to a question whose correct answer is yes.

9. Variable Liars

Problem 9.1. Why does D follow from N and C?

In what follows, the name of an inhabitant will be used interchangeably with the set of days on which said inhabitant tells the truth. I will also use standard set theory notation. Given inhabitants A and B, we can form \((A' \cap B')' = A \cup B\) only using rules N and C. Thus, D holds.

Problem 9.2. Is this island (where N and D hold) necessarily a Boolean island?

Yes it is. Given inhabitants A and B, we can form \((A' \cup B')' = A \cap B\) only using rules N and D. Thus, C holds.

Problem 9.3. Prove that Irving's Island is necessarily a Boolean island.

Condition I states that given inhabitants A and B, there is an inhabitant \(C = A' \cup B\). Using N, given inhabitants A and B, we have \(A'\). Then using I we can form \(A'' \cup B = A \cup B\). Hence, condition D holds. By Problem 9.2, the island is Boolean.

Problem 9.4. Does every Boolean island necessarily satisfy condition I of Irving's Island?

Yes, Boolean islands satisfy I. If A and B are inhabitants, then there is an inhabitant \(A'\) by N, and therefore an inhabitant \(A' \cup B\) by D. Hence, I is satisfied.

Problem 9.5. Is Edward's Island necessarily a Boolean island?

No it is not. Consider the island with a single inhabitant who always tells the truth. Then C and D are both clearly satisfied. Moreover, \(A \cup A' = A \cup \emptyset = A\), so E is satisfied. Finally, \(A = (A \cap A) \cup (A' \cap A')\), so I is satisfied.

Problem 9.6 Does a Boolean island necessarily satisfy condition E?

Yes it does, given A and B, \(C = (A \cap B) \cup (A' \cap B')\) is a native that tells the truth exactly when A and B behave the same way (this was used in my solution to the previous problem).

Problem 9.7. Which, if any, of the conditions T, F must necessarily hold on a Boolean island?

Technically neither of them need to hold. An island with no inhabitants vacuously satisfies conditions N and C, and is thus a Boolean island. However, there is not at least one inhabitant who always tells the truth or who always lies.

If we ignore the case where the island is uninhabited, then both T and F must hold. Indeed, if A is an inhabitant, \(A \cup A'\) is always truthful, and \(A \cap A'\) always lies.

Problem 9.8. Suppose an island satisfies conditions I and T. Is it necessarily a Boolean island?

No it is not. I gave an example of such an island in my solution of Problem 9.5.

Problem 9.9. What about an island satisfying conditions I and F; is it necessarily a Boolean island?

Yes, such an island is Boolean. We just need to prove that N is satisfied. The result then follows from Problem 9.3. Let A be an islander. By F, \(\emptyset\) is also an islander. By I, \(A' \cup \emptyset = A'\) is an islander. Hence, N is satisfied.

Problem 9.10. Why does it follow? (See text for context.)

If A is an inhabitant, then A and A are inhabitants, so J ensures that \(A' \cap A' = A'\) is an inhabitant. Therefore N holds.

If A and B are inhabitants, \(A' \cap B'\) is as well by J. By N, \(A \cup B\) is an inhabitant, so D holds. By Problem 9.2, Jacob's island is Boolean.
Problem 9.11. Prove that Craig was right. (See text for context.)

Suppose he was wrong, then the sociologist was correctly telling the truth: the island is not Boolean and for any inhabitants $A$ and $B$, there is an inhabitant $C = A' \cup B'$. 

So the problem is reduced to showing that the following condition $S$ implies the island is Boolean: if $A$ and $B$ are natives, then so is $A' \cup B'$. Well, $N$ is clearly satisfied, since if $A$ is an inhabitant, so is $A' \cup A' = A'$. Then, $C$ is satisfied, since if $A$ and $B$ are inhabitants, so is $A \cap B = (A' \cup B')'$ by $S$ and $N$. By Problem 9.1, Solomon’s island is Boolean.

Problem 9.12. Concerning the last problem, is it possible that the sociologist was lying? (Remember that this is an island of variable liars, in which in any one day, an inhabitant either lies the entire day or tells the truth the entire day.)

He could not have been lying. For if he was, then the island would have been Boolean. But a Boolean island obviously satisfies condition $S$, so his second statement was true, a contradiction.

Problem 9.13. Does a Boolean island necessarily satisfy condition $E'$?

The condition can be formulated as follows: if $A$ and $B$ are islanders, then so is $(A \cap B') \cup (A' \cap B) = A \Delta B$. Clearly, a Boolean island satisfies $E'$.

Problem 9.14. Someone once conjectured that if an island satisfies the condition $N$, then the conditions $E$ and $E'$ are equivalent—each implies the other. Was this conjecture correct.

Condition $E$ states that if $A$ and $B$ are natives, then so is $(A \cap B) \cup (A' \cap B')$. Condition $E'$ states that if $A$ and $B$ are natives, then so is $A \Delta B$. Note that $(A \Delta B)' = (A \cap B) \cup (A' \cap B')$, so if $N$ holds, conditions $E$ and $E'$ are equivalent. Thus, the conjecture is correct.

Problem 9.15. Suppose that an island satisfies condition $E'$ as well as condition $I$ of Irving’s island. Is such an island necessarily a Boolean island?

Yes this is true. If condition $E'$ holds, then so does $F$, since $A \Delta A = \emptyset$. By Problem 9.9, the island is Boolean.

Problem 9.16.

(a) Show that an island obeys condition $N$ then it obeys $I$ if and only if it obeys $I'$.

Condition $I$ states that if $A, B$ are islanders, then so is $A' \cup B$. Condition $I'$ states that if $A, B$ are islanders, then so is $A \cap B'$. Note that $(A' \cup B)' = A \cap B'$. Thus, if $N$ holds, $I$ and $I'$ are equivalent.

(b) Show that every Boolean island satisfies condition $I'$.

This is obvious from the way I stated condition $I'$ in part (a).

(c) Show that any island obeying conditions $E$ and $I'$ must be a Boolean island.

Let $A$ be a native. Then so is $A \cap A' = \emptyset$ by condition $I'$. Hence, $F$ holds. Then, from $E$, there is an inhabitant $(A \cap \emptyset) \cup (A' \cap \emptyset') = A'$. Hence $N$ holds. Hence, $I$ holds by part (a). By problem 9.9, the island is Boolean.