• *Part I* •

YOU MIGHT BE SURPRISED!

· 1 ·

A Diabolical Puzzle

I BELIEVE that the following puzzle may well be the most diabolical puzzle ever invented (and if it is, I proudly take credit for the invention).

Two people—A and B—each make an offer, which is given below. The problem is to determine which of the offers is better.

A's Offer: You are to make a statement. If the statement is true, you get exactly ten dollars. If the statement is false, then you get either less than ten or more than ten dollars, but not exactly ten dollars.

B's Offer: You are to make a statement. Regardless of whether the statement is true or false, you get more than ten dollars.

Which of the two offers would you prefer? Most people decide that B's offer is better, since it *guarantees* more than ten dollars, whereas with A's offer, there is no certainty of winning more than ten. And it does *seem* that B's offer is better, but seeming is sometimes deceptive. In fact, I will make an offer of my own: If any of you are willing to make me A's offer, I'll pay you twenty dollars in advance. Anyone game? (Before taking me up on it, you had better read the rest of this chapter!)

I will not give the solution to the above problem until after first considering some simpler, related puzzles.

In my book *To Mock a Mockingbird,* I presented the following puzzle: Suppose I offer two prizes—Prize 1 and Prize 2. You are to make a statement. If the statement is true, then I am to give you one of the two prizes (not saying which one). If your statement is false, then you get no prize. Obviously you can be sure of winning one of the two prizes by saying: "Two plus two is four," but suppose you have your heart set on Prize 1; what statement could you make that would guarantee that you will get Prize 1?

The solution I gave is that you say: "You will not give me Prize 2." If the statement is false, then what it says is not the case, which means that I *will* give you Prize 2. But I can't give you a prize for making a false statement, and so the statement can't be false. Therefore, it must be true. Since it is true, then what it says *is* the case, which means that you will not get Prize 2. But since your statement was true, I must give you one of the two prizes, and since it is not Prize 2, it must be Prize 1.

As we will see shortly, this little puzzle is closely related to Gödel's famous Incompleteness Theorem. To find out how, let us consider a similar puzzle (actually, the same puzzle in a different guise). First, we go to the Island of Knights and Knaves (which plays a prominent role in this book) in which every inhabitant is either a knight or a knave. On this island knights make only true statements and knaves make only false ones.

Suppose now that there are two clubs on this island—Club I and Club II. Only knights are allowed to be members of either club; knaves are rigorously excluded from both. Also, every knight is a member of one and only one of the two clubs. You visit the island one day and meet an unknown native who makes a statement from which you can deduce that he must be a member of Club I. What statement can he have made from which you can deduce this?

In analogy with the last problem, the speaker could have said: "I am not a member of Club II." If the speaker were a knave, then he really couldn't be a member of Club II and would have made a true statement, which a knave cannot do. Therefore he must be a knight, and his statement must have been true; he really isn't a member of Club II. But since he is a knight, he must be a member of a club —so he belongs to Club I.

The analogy should be obvious: Club I corresponds to those who make true statements and will get Prize 1; Club II corresponds to those who make true statements and will get Prize 2.

These puzzles embody the essential idea behind Gödel's famous sentence that asserts its nonprovability in a given mathematical system. Suppose we classify all the *true* sentences of the system (like the knights in the puzzle above) into two groups: Group I consists of all sentences of the system which, though true, are not *provable* in the system; and Group II consists of all the sentences which are not only true but are actually provable in the system. What Gödel did was to construct a sentence that asserted that it was in Group II—the sentence can be paraphrased, "I am not provable in the system." If the sentence were false, then what it says would not be the case, which would mean that it *is* provable in the system, which it cannot be (since all sentences provable in the system are true). Hence the sentence must be true, and, as it says, it is not provable in the system. Thus Gödel's sentence is true, but not provable in the system.

We will have much to say about Gödel sentences in the course of this book. For now, I wish to consider some more variants of the Prize Puzzle.

1 · First Variant

Again I offer two prizes—Prize 1 and Prize 2. If you make a true statement, I will give you at least one of the two prizes and possibly both. If you make a false statement, you get no prize. Suppose you are ambitious and wish to win both prizes. What statement would you make? (Solutions appear at the ends of chapters unless otherwise specified.)

2 • Second Variant

This time the rules change a bit: If you make a true statement, you get Prize 2; if you make a false statement, you don't get Prize 2 (you might or might not get Prize 1). Now what statement will win you Prize 1?

3 • A Perverse Variant

Suppose that in a perverse framc of mind, I now tell you that if you make a false statement, you will get one of the two prizes, but if you make a true statement, you get no prize. What statement will win Prize 1?

4 • More on the Diabolical Puzzle

Now let us return to the puzzle that opened this chapter: What was diabolical was *my* offer to pay you twenty dollars for making me A's offer, because if you took me up on it, I could win as much from you as I liked—say, a million dollars. Can you see how?

SOLUTIONS

1 • A statement that works is: "I will either get both prizes or no prize." If the statement is false, then what it says is not the case, which means that you will get exactly one prize. But you can't get a prize for a false statement. Therefore the statement must be true, and you really will get either both prizes or no prize. Since you did not make a false statement, which would result in no prize, you must get both prizes.

2 · Just say: "I will get no prize." If the statement is true, then on the one hand you get no prize (as the statement says), but on the other hand, you get Prize 2 for having made a true statement. This is a clear contradiction, hence the statement must be false. Then, unlike what the statement says, you must get a prize. You can't get Prize 2 for a false statement, so you get Prize 1.

3 · This time say: "I will get Prize 2." I leave the proof to the reader.

4 · All I have to do is say: "You will neither pay me exactly ten dollars nor exactly one million dollars." If my statement is true, then on the one hand you will not pay me either exactly ten dollars or exactly a million dollars, but on the other hand you *must* pay me exactly ten dollars for having made a true statement. This is a contradiction, hence the statement can't be true and must be false. Since it is false, what it says is not the case, which means that you *will* pay me either exactly ten dollars or exactly a million dollars, but you can't pay me exactly ten dollars for a false statement, hence you must pay me a million dollars.

Anyone still game?

• 2 •

Surprised?

LET US now turn to the paradox at the surprise examination—a paradox that is quite relevant to the consideration of Gödel in this book. We will look at it in the following form: On a Monday morning, a professor said to his class, "I will give you a surprise examination someday this week. It may be today, tomorrow, Wednesday, Thursday, or Friday at the latest. On the morning of the day of the examination, when you come to class, you will not know that this is the day of the examination."

Well, a logic student reasoned as follows: "Obviously I can't get the exam on the last day, Friday, because if I haven't gotten the exam by the end of Thursday's class, then on Friday morning I'll know that this is the day, and the exam won't be a surprise. This rules out Friday, so I now know that Thursday is the last possible day. And, if I don't get the exam by the end of Wednesday, then I'll know on Thursday morning that this must be the day (because I have already ruled out Friday), hence it won't be a surprise. So Thursday is also ruled out."

The student then ruled out Wednesday by the same argument, then Tuesday, and finally Monday, the day on which the professor was speaking. He concluded: "Therefore I can't get the exam at all; the professor cannot possibly fulfill his statement." Just then, the professor said: "Now I will give you your exam." The student was most surprised!

What was wrong with the student's reasoning?

Dozens of articles have been written about this famous problem, and uniform agreement about the correct analysis does not yet seem to have been reached. My own view is very briefly as follows:

Let me put myself in the student's place. I claim that I could get a surprise examination on any day, *even on Friday!* Here is my reasoning: Suppose Friday morning comes and I haven't gotten the exam yet. What would I then believe? Assuming I believed the professor in the first place (and this assumption is necessary for the problem), could I consistently continue to believe the professor on Friday morning if I hadn't gotten the exam yet? I don't see how I could. I could certainly believe that I would get the exam today (Friday), but I couldn't believe that I'd get a *surprise* exam today. Therefore, how could I trust the professor's accuracy? Having doubts about the professor, I wouldn't know what to believe. Anything could happen as far as I'm concerned, and so it might well be that I *could* be surprised by getting the exam on Friday.

Actually, the professor said two things: (1) You will get an exam someday this week; (2) You won't know on the morning of the exam that this is the day. I believe it is important that these two statements should be separated. It *could* be that the professor was right in the first statement and wrong in the second. On Friday morning, I couldn't consistently believe that the professor was right about both statements, but I could consistently believe his first statement. However, if I do, then his second statement is wrong (since I will then believe that I *will* get the exam today). On the other hand, if I doubt the professor's first statement, then I won't know whether or not I'll get the exam today, which means that the professor's second statement is fulfilled (assuming he keeps his word and gives me the exam). So the surprising thing is that the professor's second statement is true or false depending respectively on whether I do not or do believe his first statement. Thus the one and only way the professor can be right is if I have doubts about him; if I doubt him, that makes him right, whereas if I fully trust him, that makes him wrong! I don't know whether this curious point has been taken into consideration before.

The One-Day Version. The complication of having more than one day is really irrelevant to the heart of the problem, hence the following "one-day" version of the paradox has been proposed: The professor says to the student: "I will give you a surprise examination today." What is the student to make of that?

An equivalent problem is this: Suppose a student asks his theology professor, "Does God really exist?" The professor answers, "God exists, but you will never believe that God exists." What is the student to make of that? We will see later that under certain very reasonable assumptions about the student's reasoning ability, he cannot believe the professor without becoming inconsistent.

A problem even more pertinent to the heart of this book is this: Again, the student asks his theology professor whether God exists. This time the professor gives the following curious answer: "God exists if and only if you never believe that God exists." Equivalently, the professor is saying: "If God exists, then you will never believe that God exists, but if God doesn't exist, then you will believe that God does exist." What is the poor student to make of *that?* Can the student believe the professor without becoming inconsistent? Yes; it turns out that he can, but (again under certain reasonable assumptions about the student's reasoning ability, which will be explained in the course of this book) the student, who believes the professor, can remain consistent only if he does not *know* that he is consistent! In other words, if the student believes the professor and also believes in his own consistency, then he will become inconsistent.

This paradox is closely related to Gödel's remarkable Second Incompleteness Theorem—the theorem on the nonprovability of consistency. Gödel considered some of the most comprehensive mathematical systems known to this day. These systems are certainly consistent (since everything provable in them is true), but the amazing thing is that these systems, despite their power (or, looked at another way, because of their power), are unable to prove their own consistency. We know the consistency of these systems only by methods that cannot be formalized in the systems themselves.

This book will investigate these paradoxes and Gödel's work. To help us in this pursuit (and to have some fun!), let's turn first to some logic puzzles and their relation to propositional logic.

• Part II •

THE LOGIC OF LYING AND TRUTH TELLING

• 3 •

The Census Taker

MUCH OF the action in this book will take place on the Island of Knights and Knaves, where, as we have seen, knights always make true statements, knaves always make false statements, and every inhabitant is either a knight or a knave.

A fundamental fact about this island is that it is impossible for any inhabitant to claim to be a knave, because a knight would never lie and say he is a knave, and a knave would never truthfully admit to being a knave.

The following four problems will introduce the logical connectives *and, or, if-then,* and *if-and-only-if,* which will be dealt with more formally in Chapter 6.

McGREGOR'S VISIT

The census taker Mr. McGregor once did some fieldwork on the Island of Knights and Knaves. On this island, women are also called knights and knaves. McGregor decided on this visit to interview married couples only.

1 • (And)

McGregor knocked on one door; the husband partly opened it and asked McGregor his business. "I am a census taker," replied Mc-

Gregor, "and I need information about you and your wife. Which, if either, is a knight, and which, if either, is a knave?"

"We are both knaves!" said the husband angrily as he slammed the door.

What type is the husband and what type is the wife? (Solution follows Problem 2.)

2 · (Or)

At the next house, McGregor asked the husband: "Are both of you knaves?" The husband replied: "At least one of us is."

What type is each?

Solution to Problem 1. If the husband were a knight, he would never have claimed that he and his wife were both knaves. Therefore he must be a knave. Since he is a knave, his statement is false; so they are not both knaves. This means his wife must be a knight. Therefore he is a knave and she is a knight.

Solution to Problem 2. If the husband were a knave, then it would be true that at least one of the two is a knave, hence a knave would have made a true statement, which cannot be. Therefore the husband must be a knight. It then follows that his statement was true, which means that either he or his wife is a knave. Since he isn't a knave, then his wife is. And so the answer is the opposite of Problem 1—he is a knight and she is a knave.

The next problem is more startling than the preceding two (at least to those who haven't seen it before). It contains a theme that runs through some of the advanced problems that will crop up in later chapters.

3 · (If-Then)

The next home visited by McGregor proved more of a puzzler. The door was opened timidly by a rather shy man. After McGregor asked him to say something about himself and his wife, all the husband said was: "If I am a knight, then so is my wife."

McGregor walked away none too pleased. "How can I tell anything about either of them from such a noncommittal response?" he thought. He was about to write down "Husband and wife both unknown," when he suddenly recalled an old logic lesson from his Oxford undergraduate days. "Of course," he realized, "I can tell *both* their types!"

What type is the husband and what type is the wife?

Solution. Suppose the husband is a knight. Then it is true what he said—namely, that if he is a knight, so is his wife—and hence his wife must also be a knight. This proves that *if* the husband is a knight, so is his wife. Well, that's exactly what the husband said; he said that *if* he is a knight, then so is his wife. Therefore he made a true statement and so he must be a knight. We now know that he is a knight, and we have already proved that if he is a knight, so is his wife. Therefore the husband and wife must both be knights.

The idea behind the last problem has more far-reaching ramifications than the reader might realize. Let us consider the following variant of the problem: Suppose you visit the island prospecting for gold. Before you start digging, you want to find out whether there really is any gold on the island. It is to be assumed that each native knows whether there is any gold on the island or not. Suppose a native says to you: "If I am a knight, then there is gold on the island." You can then justifiably conclude that the native must be a knight and that there must be gold on the island. The reasoning is the same as that of the solution to Problem 3: Suppose the native is a knight, then it is really true that if he is a knight, there is gold on the island and hence that there is gold on the island. This proves that if he is a knight, then there is gold on the island. Since he said just that, he is a knight. Hence there is gold on the island.

The solution to Problem 3 and its variant are special cases of the following fact, which is sufficiently important to record as Theorem I.

Theorem I. Given any proposition p, suppose a native of the knight-knave island says: "If I am a knight, then p." Then the native must be a knight and p must be true.

The solution to Problem 3 is a special case of Theorem I, taking p to be the proposition that the native's wife is a knight. The variant of Problem 3 (the problem about the gold) is also a special case of Theorem I, taking p to be the proposition that there is gold on the island.

It also follows from Theorem I that no inhabitant of the knight-knave island can say: "If I'm a knight, then Santa Claus exists" (unless, of course, Santa Claus really does exist).

4 • (If-and-Only-If)

When the census taker interviewed the fourth couple, the husband said: "My wife and I are of the same type; we are either both knights or both knaves."

(The husband could have alternatively said: "I am a knight if and only if my wife is a knight." It comes to the same thing.)

What can be deduced about the husband and what can be deduced about the wife?

Solution. It cannot be determined whether the husband is a knight or a knave, but the wife's type can be determined as follows:

If the wife were a knave, the husband could never claim that he is the same type as his wife, because that would be tantamount to claiming that he is a knave, which he cannot do.

An alternative way of looking at the problem is this: The husband is either a knight or a knave. If he is a knight, his statement is true, hence he and his wife really are of the same type, which means his wife is also a knight. On the other hand, if he is a knave, then his statement is false, hence he and his wife are of different types, which means that his wife, unlike her husband, is a knight. And so regardless of whether the husband is a knight or a knave, his wife must be a knight. (The husband's type here is "indeterminate"; he could be a knight who truthfully claimed to be like his wife, or he could be a knave who falsely claimed to be like his wife.)

This problem is a special case of the following: Given any proposition p, suppose an inhabitant of the island says, "I am a knight if and only if p is true." What can be deduced?

Two propositions are called *equivalent* if they are either both true or both false—in other words, if either one of them is true, so is the other. Two propositions are called *inequivalent* if they are not equivalent—in other words, if one of them is true and the other is false. Now, the inhabitant has said: "I am a knight if and only if p is true." If we let k be the proposition that the inhabitant is a knight, then the inhabitant is claiming that k is equivalent to p. If he is a knight, then his claim is true, hence k really is equivalent to p; and since k is true (he is a knight), then p is also true. On the other hand, if he is a knave, then his claim is false; k is not really equivalent to p, but since k is false (he is not a knight), then again p must be true (because any proposition inequivalent to a false proposition is obviously true). And so we see that p must be true, but k is indeterminate. Let us record this as Theorem II.

Theorem II. Given any proposition p, suppose an inhabitant says: "I am a knight if and only if p." Then p must be true, regardless of whether the inhabitant is a knight or a knave.

Let us return to the problem of whether there is gold on the island. Suppose a native says: "I am a knight if and only if there is gold on the island." Then according to Theorem II (taking p to be the proposition that there is gold on the island), we see that there must be gold on the island, though we cannot determine whether the native is a knight or a knave.

We see, therefore, that if a native says: "If I am a knight, then there is gold on the island," according to Theorem I, we can infer *both* that he is a knight and that there is gold on the island. But if he instead says: "I am a knight if and only if there is gold on the island," then, according to Theorem II, all we can infer is that there is gold; we cannot determine whether the speaker is a knight or a knave.

Theorem II is the basis of a famous puzzle invented by the philosopher Nelson Goodman. The puzzle can be rendered as follows: Suppose you go to the Island of Knights and Knaves and wish to find out whether or not there is gold on the island. You meet a native and you are allowed to ask him only one question, which must be answerable by yes or no. What question would you ask him?

A question that works is: "Is it the case that you are a knight if and only if there is gold on the island?" If he answers yes, then according to Theorem II there is gold on the island. If he answers no, then there is no gold on the island (because he is denying that his being a knight is equivalent to there being gold on the island), which is tantamount to claiming that his being a knight is equivalent to there *not* being gold on the island, so again according to Theorem II there is no gold on the island.

SOME RELATED PROBLEMS

5

What statement could a native make from which you could deduce that if he is a knight, then there is gold on the island, but if he is a knave, then there might or might not be gold on the island?

6

What statement could a native make such that you could deduce that if there is gold on the island, then he must be a knight, but if there is no gold on the island, then he could be either a knight or a knave?

7

I once visited this island and asked a native: "Is there gold on this island?" All he said in reply was: "I have never claimed that there is gold on this island." Later, I found out that there *was* gold on the island. Was the native a knight or a knave?

SOLUTIONS

5 • There are many statements that would work. One such statement is: "I am a knight and there is gold on the island." Another is: "There is gold on the island and there is silver on the island." (If the native is a knight, then of course there is gold—as well as silver—but if there is gold, the native needn't be a knight—there might not be any silver.)

6 • One statement that works is: "Either I am a knight or there is gold on the island." The phrase "either–or" *means at least one—and possibly both.* And so if there is gold on the island, then it is certainly true that *either* the native is a knight *or* there is gold on the island. Therefore, if there is gold on the island, then the native's statement was true, which in turn implies that the native must be a knight. This proves that if there is gold on the island, then the native is a knight.

On the other hand, the native could be a knight without there being any gold on the island, because if he is a knight then it is true that *either* he is a knight *or* there is gold on the island.

Another statement that works is: "Either there is gold on the island or silver on the island."

7 • Suppose the native were a knave. Then his statement is false, which means that once he *did* claim that there is gold on the island. His claim must have been false (since he is a knave), which means that there is no gold on the island. But I told you that there *was* gold on the island. Hence he can't be a knave; he must be a knight.

• 4 •

In Search of Oona

THERE IS a whole cluster of knight-knave islands in the South Pacific on which some of the inhabitants are half human and half bird. These bird-people fly just as well as birds and speak as fluently as humans.

This is the story of a philosopher—a logician, in fact—who visited this cluster of islands and fell in love with a bird-girl named Oona. They were married. His marriage was a happy one, except that his wife was too flighty! For example, he would come home at night for dinner, but if it was a particularly lovely evening, Oona would have flown off to another island. So he would have to paddle around in his canoe from one island to another until he found Oona and brought her home. Whenever Oona landed on an island, the natives would all see her in the air and know she was landing. Once she was down, however, it would be very difficult to find her, so the first thing the husband did when he arrived on an island was to try to find out from the natives whether Oona had landed. What made it so difficult, of course, was that some of the natives were knaves and wouldn't tell the truth. Here are some of the incidents that befell him.

1

On one occasion, the husband came to an island in search of Oona and met two natives, A and B. He asked them whether Oona had landed on the island. He got the following responses:

A: If B and I are both knights, then Oona is on this island.
B: If A and I are both knights, then Oona is on this island.

Is Oona on this island?

2

On another occasion, two natives A and B made the following statements:

A: If either of us is a knight, then Oona is on this island.
B: That is true.

Is Oona on this island?

3

I do not remember the details of the next incident too clearly. I know that the logician met two natives A and B and that A said: "B is a knight, and Oona is on this island." But I do not remember exactly what B said. He either said: "A is a knave, and Oona is not on this island." Or he said: "A is a knave, and Oona *is* on this island." I wish I could remember which! At any rate, I do remember that the logician was able to determine whether or not Oona was on the island. Was she?

4

In the next incident, the logician came to a very small island with only six inhabitants. He interrogated each one, and curiously enough, they all said the same thing: "At least one knave on this island has seen Oona land here this evening."

Did any native of this island see Oona land there that evening?

5

In another curious incident, when the husband arrived on an island looking for Oona, he met five natives A, B, C, D, and E, who all guessed his purpose and grinned at meeting him. They made the following statements:

A: Oona is on this island.
B: Oona is *not* on this island!
C: Oona was here yesterday.
D: Oona is not here today, and she was not here yesterday.
E: Either D is a knave or C is a knight.

The logician thought about this for a while, but could get nowhere.

"Won't one of you please make another statement?" the logician pleaded. At this point A said: "Either E is a knave or C is a knight."

Is Oona on this island?

SOLUTIONS

1 · I shall content myself with briefer solutions than those I have formerly given.

Suppose A and B are both knights. Then the common statement they make is true, from which it in turn follows that Oona is on the island. So if they are both knights, then Oona is on the island. This is what they said, so they are both knights. Hence Oona is on the island.

2 · If either one is a knight, then the statement he made is true, from which it in turn follows that Oona is on the island. Therefore, if either one is a knight, Oona is on the island. Thus the statement they both made is true, hence both are knights, so surely at least one is a knight. From this and the truth of the statement they made, it follows that Oona is on the island.

3 • This is an example of what I call a *metapuzzle:* You are not told what B said, but you *are* told that from what A and B said, the logician was able to determine whether Oona was on the island. (If I had not told you that, then you couldn't possibly solve the problem!)

I will first show you that if B had said: "A is a knave, and Oona is not on this island," then the logician couldn't possibly have solved the problem. So suppose that B had said that. Now, A couldn't possibly be a knight, for if he were, then B would be a knight (as A said), which would make A a knave (as B said). Therefore A is definitely a knave. But now it could either be that B is a knight and Oona isn't on the island, or that B is a knave and Oona is on the island, and there is no way to tell which. So if B had said that, the logician couldn't have known whether Oona was on the island. But we are given that the logician did know, hence B *didn't* say that. He must have said: "A is a knave and Oona *is* on this island." Now let's see what happens.

A must be a knave for the same reason as before. If Oona is on the island, we get the following contradiction: It is then true that A is a knave and Oona is on the island, hence B made a true statement, which makes B a knight. But then A made a true statement in claiming that B is a knight and Oona is on the island, contrary to the fact that A is a knave! The only way out of the contradiction is that Oona is *not* on the island. So Oona is not on this island (and, of course, A and B are both knaves).

4 • Since all six natives said the same thing, then they are either all knights or all knaves (all knights, if what they said is true, and all knaves otherwise). Suppose they were all knights. Then it would be true that at least one knave on the island had seen Oona land, but this would be impossible, since none of them are knaves. Therefore they must all be knaves. Hence what they said is false, which means that no knave on the island saw Oona land that evening. But since all the natives are knaves, then no native at all saw Oona land that evening.

5 • We will show that if A is a knave, we get a contradiction: Suppose A is a knave. Then his second statement was false, hence E must be a knight and C must be a knave. Since E is a knight, his statement is true, hence either D is a knave or C is a knight. But C isn't a knight, so D must be a knave. Hence D's statement was false, so either Oona is here today or she was here yesterday. But Oona was not here yesterday (because C said she was and C is a knave), hence she is here today. But this makes A's first statement true, contrary to the assumption that A is a knave! Then A can't be a knave; he must be a knight. Therefore A's first statement was true, so Oona is on this island.

• 5 •

An Interplanetary Tangle

ON GANYMEDE—a satellite of Jupiter—there is a club known as the Martian-Venusian Club. All the members are either from Mars or from Venus, although visitors are sometimes allowed. An earthling is unable to distinguish Martians from Venusians by their appearance. Also, earthlings cannot distinguish either Martian or Venusian males from females, since they dress alike. Logicians, however, have an advantage, since the Venusian women always tell the truth and the Venusian men always lie. The Martians are the opposite; the Martian men tell the truth and the Martian women always lie.

One day Oona and her logician-husband visited Ganymede and were told about this club. "I'll bet you *I* can tell the men from the women and the Martians from the Venusians," said the husband proudly to his wife.

"How?" asked Oona.

"We'll visit the club tonight, which is visitors' night, and I'll show you!" said the husband (whose name, by the way, was George).

1

They visited the club that night. "Now, let's see what you can do," said Oona somewhat skeptically. "That member over there. Can you tell whether he or she is male or female?" George then went over to the member and asked him or her a single question answerable by yes or no. The member answered, and George then determined whether the member was male or female, though he could not determine whether the member was from Mars or Venus.

What question could it have been?

2

"Very clever!" said Oona, after George had explained the solution. "Now, suppose that instead of wanting to find out whether the member was male or female, you had wanted to find out whether he or she was from Mars or Venus. Could you have done that by asking only one yes-no question?"

"Of course!" said George. "Don't you see how?"

Oona thought about it for a bit and suddenly saw how. How?

3

"If you are *really* clever," said Oona, "you should be able to find out in only one question whether a given member is male or female *and* where the member is from. Let's see you do both things at once using only one yes-no question!"

"Nobody is *that* clever!" replied George. Why did George say that? (This is essentially a repetition of the last puzzle of Chapter 2 of my book *To Mock a Mockingbird.*)

4

Just then a member walked by and made a statement from which George and Oona (who was by now getting the hang of things) could deduce that the member must be a Martian female. What statement could it have been?

5

The next member who walked by made a statement from which George and Oona could deduce that the member must be a Venusian female. What statement could it have been?

6

What statement could be made by either a Martian male, a Martian female, a Venusian male, or a Venusian female?

News soon spread through the club of George and Oona's cleverness in applying logic to determine the sex and/or race of various members. The owner of the club, an American entrepreneur named Fetter, came over to George and Oona's table to congratulate them. "I would like to try you on still other members," said Fetter, "and see what you can do."

7

Just then two members walked by. "Come join us," said Fetter, who introduced them as Ork and Bog. George asked them to tell him something about themselves, and the two made the following statements.

ORK: Bog is from Venus.
BOG: Ork is from Mars.
ORK: Bog is male.
BOG: Ork is female.

From this information, George and Oona could successively identify both the sex and the race of each of them. What is Ork and what is Bog?

8

"You know," said Fetter, after Ork and Bog had left, "that Martians and Venusians often intermarry, and we have several mixed couples in this club. Here is a couple approaching us now. Let's see if you can tell whether or not it is a mixed couple."

I don't remember the couple's first names, so I will simply call them A and B.

"Where are you from?" Oona asked A.

"From Mars," was the reply.

"That's not true!" said B.

Is this a mixed couple or not?

9

"Here comes another couple," said Fetter. "Again I won't tell you whether they are a mixed couple or not. Let's see if you can figure out which one is the husband."

We will call the two A and B. George asked: "Are you both from the same planet?" Here are their replies.

A: We are both from Venus.

B: That is not true!

Which one is the husband?

10

"Here is another couple," said Fetter. "Again, I won't tell you whether it is a mixed couple or not. Let's see what will happen!"

This time I happen to remember their first names—they were Jal and Tork.

"Where are you each from?" asked George.

"My spouse is from Mars," replied Tork.

"We are both from Mars," said Jal.

This enabled George and Oona to classify both of them completely. Which is the husband and which is the wife, and which planet is each of them from?

SOLUTIONS

1 · The simplest question that works is: "Are you Martian?" Suppose you get the answer yes. The speaker is either telling the truth or lying. If the former, then the speaker is really Martian, and being a truth-telling Martian, must be male. If the speaker is lying, then the speaker is really Venusian, hence is a lying Venusian, hence is again male. So in either case, a yes answer indicates that the speaker is male. A similar analysis (which I leave to the reader) shows that a no answer indicates that the speaker is female.

Of course the question "Are you Venusian?" works equally well; a yes answer then indicates that the speaker is female, and a no answer indicates male.

2 · The question, "Are you male?" works. (I leave the verification to the reader.) Alternatively the question, "Are you female?" works as well.

3 · The reason that it is impossible to design a yes-no question that will definitely determine whether a given member is male or female *and* whether the member is Martian or Venusian is that there are four possibilities for the member—a Martian male, a Martian female, a Venusian male, a Venusian female—but there are only two possible responses to the question: yes or no. And so with only two possible responses, one cannot determine which of four possibilities holds.

4 · A simple statement that would work is: "I am a Venusian male." Obviously the statement can't be true, or we would have the contra-

diction of a Venusian male making a true statement. Hence the statement is false, which means that the speaker is *not* a Venusian male. Since the statement is false, the speaker must then be a Martian female.

5 • This is a bit trickier: One statement that works is: "I am either female or Venusian." (Note: Remember that either-or means *at least one and possibly both;* it does not mean *exactly one.*)

If the statement is false, then the speaker is neither female nor Venusian, hence must be a male Martian. But a male Martian does not make false statements, and so we get a contradiction. This proves that the statement must be true, hence the speaker must be either female or Venusian and possibly both. However, if she is female, she must also be Venusian, and if Venusian, also female, because truth-telling females are Venusian and truth-telling Venusians are female. Therefore the speaker must be both Venusian and female.

Incidentally, if the speaker had made the stronger statement, "I am female *and* Venusian," it would be impossible to determine either the sex or the race of the speaker (all that could be inferred is that the speaker is not a Martian male).

6 • One such statement is, "I am either a Martian male or a Venusian female"—or, even more simply, "I always tell the truth." Any liar or truth teller could say that.

7 • Suppose Ork told the truth. Then Bog would be both male and Venusian, hence Bog must have lied. Suppose, on the other hand, that Ork lied. Then Bog is neither male nor Venusian, hence Bog must be a Martian female, so again Bog must have lied. This proves that regardless of whether Ork told the truth or not, Bog definitely lied.

Since Bog lied, then Ork is neither from Mars nor female, hence Ork must be a Venusian male. Therefore Ork also lied, which means that Bog must be a Martian female. And so the solution is that Ork is a Venusian male and Bog is a Martian female (and all four statements were lies).

8 • Since A claimed to be from Mars, then A must be male (as we saw in the solution of Problem 1), and hence B must be female. If A is truthful, then A really is from Mars, B lied, and being a lying female is also from Mars. If A lied, A is really from Venus, B told the truth, and being female is also from Venus. Therefore this is not a mixed couple; they are both from the same planet.

9 • If A's statement is true, then both are from Venus, hence A is from Venus and A must be female. Suppose A's statement is false. Then at least one of them is from Mars. If A is from Mars, A must be female (since A's statement is false). If B is from Mars, B must be male (since B's statement is true), hence again, A must be female. And so A is the wife and B is the husband.

10 • Suppose that Jal told the truth. Then the two really are both from Mars, hence Tork is from Mars and Tork's statement that Jal is from Mars was true. We thus have the impossibility of a couple from the same planet *both* telling the truth. This cannot be, hence Jal must have lied. Therefore at least one of them is from Venus.

If Jal is from Mars, then it must be that Tork is the one from Venus. But then Tork told the truth in claiming that Jal is from Mars so Tork must be female, hence Jal must be male, and we have the impossibility of a male Martian making a false statement. Therefore Jal cannot be from Mars; Jal must be from Venus. Since Jal lied and is from Venus, Jal must be male. Also, since Jal is not from Mars, Tork lied. Hence Tork is a lying female, and thus from Mars.

In summary, Jal is a Venusian male and Tork is a Martian female.

· *Part III* ·

KNIGHTS, KNAVES, AND PROPOSITIONAL LOGIC

• 6 •

A Bit of Propositional Logic

THE LIAR–truth teller puzzles of the last three chapters take on an added significance when looked at in terms of the subject known as *propositional logic* (as we will see in the next chapter). In this chapter we will go over a few of the basics—the logical connectives, truth tables, and tautologies. Readers already familiar with these concepts might pass right on to the next chapter (or perhaps just skim this one as a refresher).

THE LOGICAL CONNECTIVES

Propositional logic, like algebra, has its own symbolism, which is relatively easy to learn. In algebra, the letters x, y, z stand for unspecified numbers; in propositional logic, we use the letters p, q, r, s (sometimes with subscripts) to stand for unspecified propositions.

Propositions can be combined by using the so-called *logical connectives.* The principal ones are:

(1) ∼ (not)
(2) & (and)
(3) v (or)

(4) $\supset$ (if-then)
(5) $\equiv$ (if-and-only-if)

An explanation of these follows.

(1) Negation. For any proposition p, by ~p we mean the *opposite* or *contrary* of p. We read ~p as "it is not the case that p"; or, more briefly, "not p." The proposition ~p is called the negation of p; it is true if p is false and it is false if p is true. We can summarize these two facts in the following table, which is called the *truth table* for negation. In this table (as in all the tables that follow), we will use the letter "T" to stand for *truth* and "F" to stand for *falsehood.*

p	~p
T	F
F	T

The first line of the truth table says that if p has the value T (i.e., if p is true), then ~p has the value F. The second line says that if p has the value F, then ~p has the value T. We can also write this as follows:

~T = F
~F = T

(2) Conjunction. For any propositions p and q, the proposition that p *and* q are both true is written "p&q" (sometimes "p∧q"). We call p&q the *conjunction* of p and q. It is true if p and q are both true, but false if either one of them is false. We thus have the following four laws of conjunction:

T & T = T
T & F = F
F & T = F
F & F = F

Thus the following is the truth table for conjunction:

p	q	p&q
T	T	T
T	F	F
F	T	F
F	F	F

*(3) **Disjunction.*** For any propositions p and q, we let pvq be the proposition that at least one of the propositions p or q is true. We read pvq as "either p or q—and possibly both." (There is another sense of "or," namely, *exactly one,* but this is not the sense in which we will use the word "or." If p and q both happen to be true, the proposition pvq is taken to be true.) The proposition pvq is called the *disjunction* of p and q. Disjunction has the following truth table:

p	q	pvq
T	T	T
T	F	T
F	T	T
F	F	F

We see that pvq is false only in the fourth case—when p and q are both false.

*(4) **If-Then.*** For any propositions p and q, we write p⊃q to mean that either p is false or p and q are both true—in other words, if p is true, so is q. We sometimes read p⊃q as "if p, then q," or "p implies q," or "it is not the case that p is true and q is false." We call p⊃q the *conditional* of p and q. For the conditional, we have the following truth table.

p	q	$p \supset q$
T	T	T
T	F	F
F	T	T
F	F	T

We note that $p \supset q$ is false only in the second line—the case when p is true and q is false. This perhaps needs some explanation: $p \supset q$ is the proposition that it is *not* the case that p is true and q is false. The only way that it can be false is if it *is* the case that p is true and q is false.

(5) If-and-Only-If. Finally, we let $p \equiv q$ be the proposition that p and q are either both true or both false, or, what is the same thing, that if either one of them is true, so is the other. We read $p \equiv q$ as "p is true if and only if q is true," or "p and q are equivalent." (We recall that two propositions are called *equivalent* if they are either both true or both false.) The proposition $p \equiv q$ is sometimes called the *biconditional* of p and q. Here is its truth table.

p	q	$p \equiv q$
T	T	T
T	F	F
F	T	F
F	F	T

Parentheses. We need to use parentheses to avoid ambiguity. For example, suppose I write p&qvr. The reader cannot tell whether I mean that p is true and either q or r is true, or whether I mean that either p and q are both true or r is true. If I mean the former, I should write p&(qvr); if I mean the latter, I should write (p&q)vr.

Compound Truth Tables. By the *truth value* of a proposition is meant its truth or falsity—that is T, if p is true, and F, if p is false.

Thus the propositions 2 + 2 = 4 and London is the capital of England, though different propositions, have the same truth value —namely, T.

Consider now two propositions p and q. If we know the truth value of p and the truth value of q, then we can determine the truth values of p&q, pvq, p⊃q, and p≡q—and also the truth value of ~p (as well as the truth value of ~q). It therefore follows that given any combination of p and q (that is, any proposition expressible in terms of p and q, using the logical connectives), we can determine the truth value of this combination once we are given the truth values of p and q. For example, suppose A is the proposition (p≡(q&p))⊃(~p⊃q). Given the truth values of p and q, we can successively find the values of q&p, p≡(q&p), ~p, (~p⊃q), and finally (p≡(q&p))⊃(~p⊃q). There are four possible distributions of truth values for p and q, and in each of the four cases we can determine the truth value of A. We can do this systematically by constructing the following table:

p	q	q&p	p≡(q&p)	~p	~p⊃q	(p≡(q&p))⊃(~p⊃q)
T	T	T	T	F	T	T
T	F	F	F	F	T	T
F	T	F	T	T	T	T
F	F	F	T	T	F	F

We see then that A is true in the first three cases and false in the fourth.

Let us consider another example: Let B = (p⊃q)⊃(~q⊃~p), and let us make a truth table for B.

p	q	~p	~q	p⊃q	~q⊃~p	(p⊃q)⊃(~q⊃~p)
T	T	F	F	T	T	T
T	F	F	T	F	F	T
F	T	T	F	T	T	T
F	F	T	T	T	T	T

We see that B is true in all four cases, and is hence an example of what is called a *tautology.*

We can also construct a truth table for a combination of three propositional unknowns—p, q, and r—but now there are eight cases to consider (because there are four distributions of T's and F's to p and q, and with each of these four distributions there are two possibilities for r). For example, suppose C is the expression (p&(q⊃r))⊃(r&∼p). It would have the following truth table:

p	q	r	q⊃r	p&(q⊃r)	∼p	r&∼p	(p&(q⊃r))⊃(r&∼p)
T	T	T	T	T	F	F	F
T	T	F	F	F	F	F	F
T	F	T	T	T	F	F	F
T	F	F	T	T	F	F	F
F	T	T	T	F	T	T	F
F	T	F	F	F	T	F	F
F	F	T	T	F	T	T	F
F	F	F	T	F	T	F	F

We will see that C is false in all eight cases; it is the very opposite of a tautology and is an example of what is called a *contradiction.* There are *no* propositions p, q, and r such that (p&(q⊃r))⊃(r&∼p) is true. (We could have seen this without a truth table by using common sense. Suppose p&(q⊃r) is true. Then how could (r&∼p) be true, since ∼p is false?)

If we make a truth table for an expression in four unknowns—say, p, q, r, and s—there are sixteen cases to consider, and so the truth table will have sixteen lines. In general, for any positive whole number n, a truth table for an expression in n unknowns must have 2^n lines (each time we add an unknown, the number of lines doubles).

TAUTOLOGIES

A proposition is called a *tautology* if it can be established purely on the basis of the truth table rules for the logical connectives. For example, suppose that one person says that it will rain tomorrow and a second person says that it won't. We can hardly expect to tell which one is right by using a truth table. We must wait till tomorrow and then *observe* the weather. But suppose a third person says today: "Either it will rain tomorrow or it won't." Now, that's what I would call a *safe* prediction! Without waiting for tomorrow, and making an observation, we know *by pure reason* that he must be right. His assertion is of the form $p \vee \sim p$ (where p is the proposition that it will rain tomorrow), and for *every* proposition p, the proposition $p \vee \sim p$ must be true (as a truth table will easily show).

The more usual definition of tautology involves the notion of a *formula.* By a formula is meant any expression built from the symbols $\sim$, &, v, $\supset$, $\equiv$ and the propositional variables p, q, r, . . . parenthesized correctly. Here are the precise rules for constructing formulas:

(1) Any propositional variable standing alone is a formula.

(2) Given any formulas X and Y already constructed, the expressions (X&Y), (XvY), $(X \supset Y)$, or $(X \equiv Y)$ are again formulas, and so is the expression $\sim X$.

It is to be understood that no expression is a formula unless it is constructed according to rules (1) and (2) above.

When displaying a formula standing alone, we can dispense with the outermost parentheses without incurring any ambiguity—for example, when we say "the formula $p \supset q$," we mean "the formula $(p \supset q)$."

A formula in itself is neither true nor false, but only becomes true or false when we *interpret* the propositional variables as standing for definite propositions. For example, if I asked: "Is the formula (p&q) true?", you would probably (and rightly) reply: "It depends on what propositions the letters 'p' and 'q' represent." And so a formula such as "p&q" is sometimes true and sometimes false. On the other hand, a formula such as "pv~p" is *always* true (it is true whatever proposition is represented by the letter "p") and is accordingly called a *tautological* formula. Thus a tautological formula is by definition a formula that is *always* true—or what is the same thing, a truth table for the formula will have only T's in the last column. We can then define a *proposition* to be a tautology if it is expressed by some tautological formula under some interpretation of the propositional variables. (For example, the proposition that it is either raining or not raining is expressed by the formula pv~p, if we interpret "p" to be the proposition that it is raining.)

Logical Implication and Equivalence. Given any two propositions X and Y, we say that X *logically* implies Y, or that Y is a *logical consequence* of X if X⊃Y is a tautology. We say that X is *logically equivalent* to Y if X≡Y is a tautology; or, what is the same thing, if X logically implies Y and Y logically implies X.

SOME TAUTOLOGIES

The truth table is a systematic method of verifying tautologies, but many tautologies can be more quickly recognized by using a little common sense. Here are some examples:

(1) ((p⊃q)&(q⊃r))⊃(p⊃r).

This says that if p implies q, and if q implies r, then p implies r. This is surely self-evident (but can, of course, be verified by a truth table). This tautology has a name—it is called a *syllogism.*

(2) (p&(p⊃q))⊃q.

This says that if p is true, and if p implies q, then q is true. This is sometimes paraphrased: "Anything implied by a true proposition is true."

(3) $((p \supset q) \& {\sim} q) \supset {\sim} p$.

This says that if p implies a *false* proposition, then p must be false.

(4) $((p \supset q) \& (p \supset {\sim} q)) \supset {\sim} p$.

This says that if p implies q and also p implies not q, then p must be false.

(5) $(({\sim} p \supset q) \& ({\sim} p \supset {\sim} q)) \supset p$.

This principle is known as *reductio ad absurdum.* To show that p is true, it suffices to show that ~p implies some proposition q as well as its negation ~q.

(6) $((p \vee q) \& {\sim} p) \supset q$.

This is a familiar principle of logic: If at least one of p or q is true, and if p is false, then it must be q that is true.

(7) $((p \vee q) \& ((p \supset r) \& (q \supset r))) \supset r$.

This is another familiar principle known as *proof by cases.* Suppose pvq is true. Suppose also that p implies r and that q implies r. Then r must be true (regardless of whether it is p or q that is true —or both).

The reader with little experience in propositional logic should benefit from the following exercise.

Exercise 1. State which of the following are tautologies.

(a) $(p \supset q) \supset (q \supset p)$
(b) $(p \supset q) \supset ({\sim} p \supset {\sim} q)$
(c) $(p \supset q) \supset ({\sim} q \supset {\sim} p)$

(d) $(p \equiv q) \supset (\sim p \equiv \sim q)$
(e) $\sim(p \supset \sim p)$
(f) $\sim(p \equiv \sim p)$
(g) $\sim(p \& q) \supset (\sim p \& \sim q)$
(h) $\sim(p v q) \supset (\sim p v \sim q)$
(i) $(\sim p v \sim q) \supset \sim(p v q)$
(j) $\sim(p \& q) \equiv (\sim p v \sim q)$
(k) $\sim(p v q) \equiv (\sim p \& \sim q)$
(l) $(q \equiv r) \supset ((p \supset q) \equiv (p \supset r))$
(m) $(p \equiv (p \& q)) \equiv (q \equiv (p v q))$

Answers. (a) No, (b) No, (c) Yes, (d) Yes, (e) No! (f) Yes, (g) No, (h) Yes, (i) No, (j) Yes, (k) Yes, (l) Yes, (m) Yes (both $p \equiv (p \& q)$ and $(q \equiv (p v q))$ are equivalent to $p \supset q$).

Concerning (e), many beginners think that no proposition p can imply its negation. This is not so! If p happens to be false, then $\sim p$ is true, hence in that case, $p \supset \sim p$ *is* true. However, no proposition can be *equivalent* to its own negation, and so (f) is indeed a tautology.

Discussion. The significance of tautologies is that they are not only true, but *logically certain.* No scientific experiments are necessary to establish their truth—they can be verified on the basis of *pure reason.*

One can alternatively characterize tautologies without appeal to the notion of formulas. Let us define a *state of affairs* as any classification of all propositions into two categories—*true* propositions and *false* propositions—subject to the restriction that the classification must obey the truth table conditions for the logical connectives (for example, we may not classify pvq as true if p and q are both classified as false). A tautology, then, is a proposition which is true in *every possible state of affairs.*

This is related to Leibniz's notion of other possible worlds. Leibniz claimed that of all possible worlds, this one was the best. Frankly, I have no idea whether he was right or wrong in this, but the interesting thing is that he considered other possible worlds. Out of this, a whole branch of philosophical logic known as possible world semantics has developed in recent years—notably by the philosopher Saul Kripke—which we will discuss in a later chapter. Given any possible world, the set of all propositions that are true for that world, together with the set of all propositions that are false for that world, constitute the state of affairs holding for *that* world. A tautology, then, is true, not only for this world, but for *all* possible worlds. The physical sciences are interested in the state of affairs that holds for the *actual* world, whereas pure mathematics and logic study *all* possible states of affairs.

• 7 •

Knights, Knaves, and Propositional Logic

KNIGHTS AND KNAVES REVISITED

We can now introduce a simple but basic translation device whereby a host of problems about liars and truth tellers can be reduced to problems in propositional logic. This device will be crucial in several subsequent chapters.

Let us return to the Island of Knights and Knaves. Given a native P, let k be the proposition that P is a knight. Now, suppose that P asserts a proposition X. In general we do not know whether P is a knight or a knave, nor whether X is true or false. But this much we do know: If P is a knight, then X is true, and conversely, if X is true, then P is a knight (because knaves never make true statements). And so we know that P is a knight if and only if X is true; in other words, we know that the proposition $k \equiv X$ is a true proposition. And so we translate "P asserts X" as "$k \equiv X$."

Sometimes we have more than two natives involved—for example, suppose we have two natives P_1 and P_2. We let k_1 be the proposition that P_1 is a knight; we let k_2 be the proposition that P_2 is a knight. If a third native P_3 is involved, we let k_3 be the proposition that P_3 is a knight, and so forth, for all natives involved. We then translate "P_1 asserts X" as "$k_1 \equiv X$"; we translate "P_2 asserts X" as "$k_2 \equiv X$"; and so on.

Let us now look at the first problem in Chapter 3 (page 15). There are two natives P_1 and P_2 (the husband and the wife) involved. We are given that P_1 asserts that P_1 and P_2 are both knaves; we are to determine the types of P_1 and P_2. Now, k_1 is the proposition that P_1 is a knight, hence $\sim k_1$ is equivalent to the proposition that P_1 is a knave (since each inhabitant is a knight or a knave, but not both). Similarly, $\sim k_2$ is the proposition that P_2 is a knave. Hence the proposition that P_1 and P_2 are both knaves is $\sim k_1 \& \sim k_2$. P_1 is asserting the proposition $\sim k_1 \& \sim k_2$. Then, using our translation device, the *reality* of the situation is that $k_1 \equiv (\sim k_1 \& \sim k_2)$ is true. So the problem can be posed in the following purely propositional terms: Given two propositions k_1 and k_2 such that $k_1 \equiv (\sim k_1 \& \sim k_2)$ is true, what are the truth values of k_1 and k_2? If we make a truth table, we car hat the only case in which $k_1 \equiv (\sim k_1 \& \sim k_2)$ comes out true is when k_1 is false and k_2 is true. (We also saw this by common-sense reasoning when we solved the problem in Chapter 3.) The upshot is that the proposition $(k_1 \equiv (\sim k_1 \& \sim k_2)) \supset \sim k_1$ is a tautology, and so is the proposition $(k_1 \equiv (\sim k_1 \& \sim k_2)) \supset k_2$.

The entire mathematical content of this problem is that for any propositions k_1 and k_2, the following proposition is a tautology: $(k_1 \equiv (\sim k_1 \& \sim k_2)) \supset (\sim k_1 \& k_2)$.

The reader might note as well that the converse proposition $(\sim k_1 \& k_2) \supset (k_1 \equiv (\sim k_1 \& \sim k_2))$ is also a tautology, hence the proposition $k_1 \equiv (\sim k_1 \& \sim k_2)$ is logically *equivalent* to the proposition $\sim k_1 \& k_2$.

Now let us look at the second problem in Chapter 3. Here P_1 asserts that either P_1 or P_2 is a knave. We concluded that P_1 is a knight and P_2 is a knave. The mathematical content of this fact is that the proposition $(k_1 \equiv (\sim k_1 \vee \sim k_2)) \supset (k_1 \& \sim k_2)$ is a tautology.

The reader might note in passing that the converse is also true, and hence that the proposition $(k_1 \equiv (\sim k_1 \vee \sim k_2)) \equiv (k_1 \& \sim k_2)$ is a tautology.

The translation of Problem 3 is of particular theoretical significance; in fact, let us consider it in the more general form of Theorem I, Chapter 3 (page 18). We have a native P claiming about a certain proposition q that if P is a knight, then q is true (q could be the proposition that P's wife is a knight, or that there is gold on the island, or any proposition whatsoever). We let k be the proposition that P is a knight. Thus P is claiming the proposition $k \supset q$, and so the reality of the situation is that $k \equiv (k \supset q)$ is true. From this we are to determine the truth value of k and q. As we have seen, k and q must both be true. Thus the mathematical content of Theorem I, Chapter 3, is that $(k \equiv (k \supset q)) \supset (k \& q)$ is a tautology. Of course this fact is not really dependent on the particular nature of the proposition k; for *any* proposition p and q, the proposition $(p \equiv (p \supset q)) \supset (p \& q)$ is a tautology. The converse proposition, $(p \& q) \supset (p \equiv (p \supset q))$, is also a tautology, because if $p \& q$ is true, p and q are both true, then $p \supset q$ must be true, hence $p \equiv (p \supset q)$ must be true. And so the following is a tautology: $(p \equiv (p \supset q)) \equiv (p \& q)$.

Let us now look at Problem 4, or rather Theorem II, in Chapter 3. Here we have P claiming that P is a knight *if and only if* q. Again we let k be the proposition that P is a knight, and so P is claiming the proposition $k \equiv q$. Therefore we know that $k \equiv (k \equiv q)$ is true. From this we can determine that q must be true, and so the essential mathematical content of Theorem II, Chapter 3, is that the following is a tautology: $(k \equiv (k \equiv q)) \supset q$.

This tautology is what I would call the Goodman tautology, since it arises from Nelson Goodman's problem, which is discussed in Chapter 3.

Exercise 1. Let us consider three inhabitants P_1, P_2, and P_3 of the knight-knave island. Suppose P_1 and P_2 make the following statements.

P_1: P_2 and P_3 are both knights.
P_2: P_1 is a knave and P_3 is a knight.

What types are P_1, P_2, and P_3?

Exercise 2. (a) Is the following a tautology?

$$((k_1 \equiv (k_2 \& k_3)) \& (k_2 \equiv (\sim k_1 \& k_3))) \supset ((\sim k_1 \& \sim k_2) \& \sim k_3$$

(b) How does this relate to Exercise 1?

Exercise 3. Show that the proposition p_3 is a logical consequence of the following two propositions:

(1) $p_1 \equiv \sim p_2$
(2) $p_2 \equiv (p_1 \equiv \sim p_3)$

Exercise 4. Suppose P_1, P_2, and P_3 are three inhabitants of the knight-knave island, and that P_1 and P_2 make the following statements:

P_1: P_2 is a knave.
P_2: P_1 and P_3 are of different types.

(a) Is P_3 a knight or a knave?
(b) How does this relate to Exercise 3?

THE OONA PROBLEMS

Many of the Oona problems of Chapter 4 can also be solved by truth tables. Consider, for example, the first one (page 23): We have two natives P_1 and P_2; P_1 asserts that if P_1 and P_2 are both knights, then Oona is on the island. P_2 asserts the same thing. We let O be the proposition that Oona is on the island. Using our translation device, we know that the following two propositions are true.

(1) $k_1 \equiv ((k_1 \& k_2) \supset O)$
(2) $k_2 \equiv ((k_1 \& k_2) \supset O)$

We are to determine whether O is true or false. Well, if you make a truth table for the conjunction of (1) and (2)—i.e., for $(k_1 \equiv ((k_1 \& k_2) \supset O)) \& (k_2 \equiv ((k_1 \& k_2) \supset O))$, you will see that the last column contains T's in those and only those places in which O has the value T. Therefore O must be true.

Exercise 5. Suppose the husband goes looking for Oona and meets two natives A and B on an island and they make the following statements:

A: If B is a knight, then Oona is not on this island.
B: If A is a knave, then Oona is not on this island.

Is Oona on this island?

Exercise 6. On another island, two natives A and B make the following statements:

A: If either of us is a knight, then Oona is on this island.
B: If either of us is a knave, then Oona is on this island.

Is Oona on this island?

Exercise 7. On another island, two natives A and B make the following statements.

A: If I am a knight and B is a knave, then Oona is on this island.
B: That is not true!

Is Oona on this island?

MARTIANS AND VENUSIANS REVISITED

Many of the Mars–Venus–Female–Male puzzles of Chapter 5 can also be solved by truth tables, although the translation device needed is a bit more complex, and this type of problem is not considered further in the present book.

Let us number the club members in some order P_1, P_2, P_3, etc., and for each number i, let V_i be the proposition that P_i is Venusian, and let F_i be the proposition that P_i is female. Then P_i is *Martian* can be written $\sim V_i$, and P_i is *male* can be written $\sim F_i$. Now, P_i tells the truth if and only if P_i is either a Venusian female or a Martian male, which can be symbolized $(V_i \& F_i) v (\sim V_i \& \sim F_i)$, or more simply, $V_i \equiv F_i$. And so now if P_i asserts a proposition X, the reality of the situation is the following proposition: $(V_i \equiv F_i) \equiv X$.

This then is our translation device. Whenever P_i asserts X, we write down: $(V_i \equiv F_i) \equiv X$.

Consider, for example, Problem 7 of Chapter 5 (page 30)—the case of Ork and Bog. Let P_1 be Ork and P_2 be Bog. Then we are given the following four propositions (after we apply the translation device):

(1) $(V_1 \equiv F_1) \equiv V_2$
(2) $(V_2 \equiv F_2) \equiv \sim V_1$
(3) $(V_1 \equiv F_1) \equiv \sim F_2$
(4) $(V_2 \equiv F_2) \equiv F_1$

The truth values of V_1, F_1, V_2, and F_2 can then be found by a truth table. (There are four unknowns V_1, F_1, V_2, and F_2 involved, and so there are sixteen cases to consider!)

Note: Not all liar–truth teller problems can be solved by the translation devices in this chapter. These devices work fine for problems in which we are told what the speakers say and must then deduce certain facts about them. But for the more difficult type of puzzles in which we are to design a question or a statement to do a certain job, more thought is needed. There are other systematic devices that help in many cases, but this is a topic outside the main line of thought in this book.

ANSWERS TO EXERCISES

1 · All three are knaves.

2 · The proposition is a tautology.

3 · We leave this to the reader.

4 · P_3 is a knight.

5 · Oona is not on the island (and both natives are knights).

6 · Oona is on the island (and both natives are knights).

7 · Oona is on the island (and A is a knight; B is a knave).

· 8 ·

Logical Closure and Consistency

INTERDEFINABILITY OF THE LOGICAL CONNECTIVES

We are now familiar with the five basic logical connectives: $\sim$, &, v, $\supset$, and $\equiv$. We could have started with fewer and defined the others in terms of them. This can be done in several ways, some of which will be illustrated in the following puzzles.

1

Suppose an intelligent man from Mars comes down to Earth and wants to learn our logic. He claims to understand the words "not" and "and," but he does not know the meaning of the word "or." How could we explain it to him, using only the notions of *not* and *and?*

Let me rephrase the problem. Given any proposition p, he knows the meaning of $\sim$p, and given any propositions p and q, he knows the meaning of p&q. What the Martian is looking for is a way of writing down an expression or formula in the letters p and q, using *only* the logical connectives $\sim$ and &, such that the expression is logically equivalent to the expression pvq. How can this be done?

2

Now suppose a lady from the planet Venus comes here and tells us that she understands the meanings of $\sim$ and v, but wants an explanation of &. How can we define & in terms of $\sim$ and v? (That is, what expression in terms of p, q, $\sim$, and v is logically equivalent to p&q?)

3

This time a being from Jupiter comes down who understands $\sim$, &, and v, and wants us to define $\supset$ in terms of these. How can this be done?

4

Next comes a being from Saturn who strangely enough understands $\sim$ and $\supset$, but does not understand either & or v. How can we explain & and v to him?

5

Now comes a being from Uranus who understands only the connective $\supset$. It is not then possible to explain to him what & means, nor what $\sim$ means, but it *is* possible to define v in terms of just the one connective $\supset$. How can this be done? (The solution is not at all obvious. That it is possible to do it is part of the folklore of mathematical logic, but I have been unable to find out the logician who discovered it.)

6

It is obvious that $\equiv$ can be defined in terms of $\sim$, &, and v. Show two different ways of doing this.

7 · A Special Problem

Suppose a being from outer space understands the meanings of ⊃ and ≡. It will not then be possible to explain to him or her the meaning of ~, but it will be possible to explain &. How can this be done? (This discovery is mine, as far as I know.)

Two Other Connectives. We now see that the logical connectives ~, &, v, ⊃, and ≡ can all be defined from just the two connectives ~ and &, or alternatively from ~ and v, or alternatively from ~ and ⊃. Is there just *one* logical connective from which all five connectives can be defined? This problem was solved in 1913 by the logician Henry M. Sheffer. He defined p|q to mean that p and q are not both true. The symbol "|" is known as the "Sheffer stroke"; we can read p|q as "p and q are incompatible" (at least one is false). He showed that ~, &, v, ⊃, ≡ are all definable from the stroke symbol. Another logical connective that gives all other connectives is ↓; this symbol is called the symbol for *joint denial.* We read p↓q as "p and q are both false," or "neither p nor q is true."

8

How can ~, &, v, ⊃, ≡ all be defined from Sheffer's stroke? How can they all be defined from joint denial?

The Logical Constant ⊥. A proposition X is called *logically contradictory* or *logically false* if its negation ~X is a tautology. For example, for any proposition p, the proposition (p&~p) is logically false. So is p≡~p.

We shall use the now standard symbol "⊥" as representing any one particular logical falsehood (which one doesn't matter). This can be regarded as fixed for the remainder of this book—and any other books you might read in which this symbol appears. (The symbol "⊥" is pronounced "eet"; it is the symbol "T" written upside down.)

For any proposition p, the proposition $\bot \supset p$ is a tautology (because $\bot$ is logically false, so $\bot \supset p$ is true, regardless of whether p is true or false). Thus *every* proposition is a logical consequence of $\bot$. (It's a good thing that $\bot$ itself isn't true, for if it were, everything would be true and the whole world would explode!)

Many modern formulations of propositional logic build their entire theory on just $\supset$ and $\bot$, because all the other logical connectives can be defined from these two (see Problem 9 below). This is the course we will adopt because it fits in best with the problems of this book. One then defines T as $\bot \supset \bot$. We refer to $\bot$ as *logical falsehood* and T as *logical truth* (obviously T is a tautology).

9

How does one define all the logical connectives from $\supset$ and $\bot$?

LOGICAL CLOSURE

A Logically Qualified Machine. To illustrate the important notion of *logical closure,* let us imagine a computing machine programmed to prove various propositions. Whenever the machine proves a proposition, it prints it out (more precisely, it prints out a *sentence* that expresses the proposition). The machine, if left to itself, will run on forever.

We will call the machine *logically qualified* if it satisfies the following two conditions:

(1) Every tautology will be proved by the machine sooner or later.

(2) For any proposition p and q, if the machine ever proves p and proves $p \supset q$, then it will sooner or later prove q. (We might think of the machine as *inferring* q from the two propositions p and $p \supset q$. Of course the inference is valid.)

Logically qualified machines have one very important property, of which the following problem illustrates some examples.

10

Suppose a machine is logically qualified.

(a) If the machine proves p, will it necessarily prove $\sim\sim p$?

(b) If the machine proves p and proves q, will it necessarily prove the single proposition p&q?

Logical Closure. There is an old rule in logic called *modus ponens,* which is that having proved p and having proved $p \supset q$, one can then infer q. A set C of propositions is said to be *closed* under modus ponens if for any propositions p and q, if p and $p \supset q$ are both in the set C, so is the proposition q.

We can now define a set C of propositions to be *logically closed* if the following two conditions hold:

Condition 1. C contains all tautologies.

Condition 2. For any propositions p and q, if p and $p \supset q$ are both in C, so is q.

Thus a logically closed set is a set that contains all tautologies and that is closed under modus ponens. (The reason for the term *logically closed* will soon be apparent.)

To say that a machine is logically qualified is tantamount to saying that the set C of all propositions that the machine can prove is a logically closed set. However, logically closed sets also arise in situations in which no machines are involved. For example, we will be considering mathematical systems in which the set of all propositions provable in the system is a logically closed set. Also, a good part of this book will be dealing with logicians whose set of beliefs is a logically closed set.

Logical Consequence. We have defined Y to be a logical consequence of X if the proposition $X \supset Y$ is a tautology. We shall say that Y is a logical consequence of two propositions X_1 and X_2 if it is a consequence of the proposition $X_1 \& X_2$—in other words, if $(X_1 \& X_2) \supset Y$ is a tautology, or, what is the same thing, if the proposition $X_1 \supset (X_2 \supset Y)$ is a tautology. We define Y to be a logical consequence of X_1, X_2, and X_3 if $((X_1 \& X_2) \& X_3) \supset Y$ is a tautology—or, what is the same thing, if $X_1 \supset (X_2 \supset (X_3 \supset Y))$ is a tautology. More generally, for any finite set of propositions $X_1, \ldots, X_n$, we can define Y to be a logical consequence of this set if the proposition $(X_1 \& \ldots \& X_n) \supset Y$ is a tautology.

The importance of logically closed sets is that they enjoy the following property:

Principle L (the Logical Closure Principle). If C is logically closed, then for any n proposition $X_1, \ldots, X_n$ in C, all logical consequences of these n propositions are also in C.

Discussion. Returning to our example of a logically qualified machine, Principle L tells us that if the machine should ever prove a proposition p, then it will sooner or later prove all logical consequences of p; if the machine should ever prove two propositions p and q, it will sooner or later prove all logical consequences of p and q—and so forth, for any finite number of propositions.

The same applies to logicians whose set of beliefs is logically closed. If a logician ever believes p, he will sooner or later believe all logical consequences of p; if he should ever believe p and q, he will believe all logical consequences of p and q; and so on.

11

Why is Principle L correct?

12

Another important property of logically closed sets is that if C is logically closed and contains some proposition p and its negation ~p, then every proposition must be in C.

Why is this so?

CONSISTENCY

The last problem brings us to the important notion of *consistency*. A logically closed set C will be called *inconsistent* if it contains ⊥ and *consistent* if it doesn't contain ⊥.

The following is another important property of logically closed sets.

Principle C. Suppose C is logically closed. Then the following three conditions are all equivalent (any one of them implies the other two):

(1) C is inconsistent (C contains ⊥).

(2) C contains *all* propositions.

(3) C contains some proposition p and its negation ~p.

Note: Given a set S of propositions that is not logically closed, S is said to be inconsistent if ⊥ is a *logical consequence* of some finite subset $X_1, \ldots, X_n$ of propositions in S. (This is, incidentally, equivalent to saying that *every* proposition is a logical consequence of some finite subset of S.) However, we will be dealing almost exclusively with sets that *are* logically closed.

13

Prove Principle C.

SOLUTIONS

1 • To say that at least one of the propositions p and q is true is to say that it is *not* the case that p and q are both false; in other words, it is not the case that ~p and ~q are both true. And so pvq is equivalent to the proposition ~(~p&~q). Since the Martian understands ~ and &, then he will understand ~(~p&~q). And so you

can say to the Martian: "When I say p or q, all I mean is that it is not the case that not p and not q."

2 • To say that p and q are *both* true is equivalent to saying that it is *not* the case that either p or q is false—in other words, it is not the case that either $\sim p$ or $\sim q$ is true. And so $p \& q$ is logically equivalent to the proposition $\sim(\sim p \vee \sim q)$.

3 • This can be done in several ways: On the one hand, $p \supset q$ is logically equivalent to $\sim p \vee (p \& q)$. It is also equivalent to $\sim(p \& \sim q)$ (it is not the case that p is true and q is false), and to $\sim p \vee q$.

4 • $p \vee q$ is logically equivalent to $\sim p \supset q$. And so we can get v in terms of $\sim$ and $\supset$. Once we have $\sim$ and v, we can get & by the solution to Problem 2 above. More directly, $p \& q$ is equivalent to $\sim(p \supset \sim q)$, as the reader can verify.

5 • This is tricky indeed! The proposition $p \vee q$ happens to be logically equivalent to $(p \supset q) \supset q$, as the reader can verify by a truth table.

6 • $p \equiv q$ is obviously equivalent to $(p \supset q) \& (q \supset p)$. It is also equivalent to $(p \& q) \vee (\sim p \& \sim q)$.

7 • We have already solved this in the last chapter, in connection with the third knight-knave problem, page 16; we showed that $p \& q$ is logically equivalent to $p \equiv (p \supset q)$.

8 • Given Sheffer's stroke, we can get the other connectives as follows: First of all, $\sim p$ is logically equivalent to $p|p$. (The proposition $p|p$ is that at least one of the propositions p or p is false, but since the two propositions p and p are the same, this simply says that p is false.) Now that we have $\sim$, we can define $p \vee q$ to be $\sim(p|q)$. (Since $p|q$ is the proposition that at least one of p or q is false, its negation $\sim(p|q)$ is equivalent to saying that they are not both false—i.e., that

at least one is true.) Once we have ~ and v, we can get & (by Problem 2), then ⊃ and ≡ (by Problems 3 and 6). We thus get ~, &, v, ⊃, and ≡ from Sheffer's stroke.

If we start with joint denial ↓, instead of Sheffer's stroke, we proceed as follows: We first take ~p to be p↓p. Then we take p&q to be ~(p↓q). Having gotten ~ and &, we can then get all the rest in the manner we have seen.

9 • The proposition ~p is logically equivalent to p⊃⊥, and so we can get ~ from ⊃ and ⊥. Once we have ~ and ⊃, we can get v and & (by Problem 4). Then we can get ≡ from ⊃ and &.

10 • (a) Suppose the machine proves p. It will also prove p⊃~~p (since this is a tautology), hence it will print ~~p (by the second condition defining logical qualification).

(b) Suppose the machine proves p and proves q. Now, the proposition p⊃(q⊃(p&q)) is a tautology, hence the machine will prove it. Once the machine has proved p and p⊃(q⊃(p&q)), it must prove q⊃(p&q). Once the machine has proved this, then, since it proves q, it must prove p&q.

(Actually this problem is but a special case of the next, as the reader will see.)

11 • Let us first consider the case $n = 1$: Suppose X_1 is in C, and Y is a logical consequence of X_1. Then $X_1 \supset Y$ is a tautology, hence is in C (by Condition 1). Since X_1 and $X_1 \supset Y$ are both in C, so is Y (by Condition 2).

Now let us consider the case $n = 2$: Suppose X_1 and X_2 are both in C, and Y is a logical consequence of X_1 and X_2. Then $(X_1 \& X_2) \supset Y$ is a tautology, hence $X_1 \supset (X_2 \supset Y)$ is a tautology (as the reader can verify) and is therefore in C. Since X_1 and $X_1 \supset (X_2 \supset Y)$ are both in C, so is $X_2 \supset Y$ (Condition 2). Since X_2 and $X_2 \supset Y$ are in C, so is Y (again by Condition 2).

For the case $n = 3$, suppose Y is a logical consequence of X_1, X_2, and X_3. Then $X_1 \supset (X_2 \supset (X_3 \supset Y))$ is a tautology—it is logically equivalent to $(X_1 \& X_2 \& X_3) \supset Y$. Then using Condition 2 three times, we successively get $X_2 \supset (X_3 \supset Y)$ in C, $(X_3 \supset Y)$ in C, and finally Y in C.

It should be obvious how this proof generalizes for any positive whole number n.

Note: We have remarked that Problem 10 is but a special case of the present problem. The reason, of course, is that $\sim\sim p$ is a logical consequence of p, hence any logically closed set containing p must also contain $\sim\sim p$. Also, p&q is a logical consequence of the two propositions p and q, so any logically closed set containing both p and q must also contain p&q.

12 • Suppose that p and its negation $\sim p$ are both in C and that C is logically closed. Let q be any proposition whatsoever. The proposition $(p \& \sim p) \supset q$ is a tautology (as the reader can check by a truth table, or more simply by observing that since $p \& \sim p$ must be false, then $(p \& \sim p) \supset q$ must be true). And so q is a logical consequence of the true propositions p and $\sim p$. Then according to Principle L, the proposition q must also be in C. And so *every* proposition q is in C.

13 • We will show that the three conditions are all equivalent by showing that (1) implies (2), which in turn implies (3), which in turn implies (1).

Suppose (1). Since every proposition is a logical consequence of $\perp$ and $\perp$ is in C, then every proposition is in C (by Principle L). Thus (2) holds.

It is completely obvious that (2) implies (3), because if *all* propositions are in C, then for *any* proposition p, both p and $\sim p$ are in C.

Now suppose that (3) holds—i.e., that for some p, both p and $\sim p$ are in C. Since $\perp$ is a logical consequence of p and $\sim p$, then $\perp$ must be in C (by Principle L)—i.e., C must be inconsistent.