CHAPTER ONE

The Paradox of the Unexpected Hanging

"A NEW AND POWERFUL PARADOX has come to light." This is the opening sentence of a mind-twisting article by Michael Scriven that appeared in the July 1951 issue of the British philosophical journal *Mind*. Scriven, who bears the title of "professor of the logic of science" at the University of Indiana, is a man whose opinions on such matters are not to be taken lightly. That the paradox is indeed powerful has been amply confirmed by the fact that more than 20 articles about it have appeared in learned journals. The authors, many of whom are distinguished philosophers, disagree sharply in their attempts to resolve the paradox. Since no consensus has been reached, the paradox is still very much a controversial topic.

No one knows who first thought of it. According to the Harvard University logician W. V. Quine, who wrote one of the articles (and who discussed paradoxes in *Scientific American* for April 1962), the paradox was first circulated by word of mouth in the early 1940s. It often took the form of a puzzle about a man condemned to be hanged.

The man was sentenced on Saturday. "The hanging will take place at noon," said the judge to the prisoner, "on one of the seven days of next week. But you will not know which day it is until you are so informed on the morning of the day of the hanging."

The judge was known to be a man who always kept his word. The prisoner, accompanied by his lawyer, went back to his cell. As soon as the two men were alone the lawyer broke into a grin. "Don't you see?" he exclaimed. "The judge's sentence cannot possibly be carried out."

"I don't see," said the prisoner.

"Let me explain. They obviously can't hang you next Saturday. Saturday is the last day of the week. On Friday afternoon you would

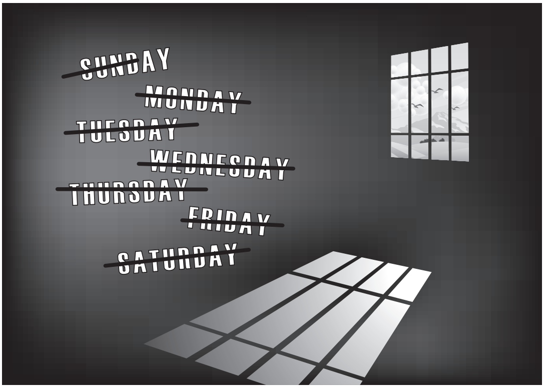

Figure 1. The prisoner eliminates all possible days.

still be alive and you would know with absolute certainty that the hanging would be on Saturday. You would know this *before* you were told so on Saturday morning. That would violate the judge's decree."

"True," said the prisoner.

"Saturday, then is positively ruled out," continued the lawyer. "This leaves Friday as the last day they can hang you. But they can't hang you on Friday because by Thursday afternoon only two days would remain: Friday and Saturday. Since Saturday is not a possible day, the hanging would have to be on Friday. Your knowledge of that fact would violate the judge's decree again. So Friday is out. This leaves Thursday as the last possible day. But Thursday is out because if you're alive Wednesday afternoon, you'll know that Thursday is to be the day."

"I get it," said the prisoner, who was beginning to feel much better. "In exactly the same way I can rule out Wednesday, Tuesday and Monday. That leaves only tomorrow. But they can't hang me tomorrow because I know it today!"

In brief, the judge's decree seems to be self-refuting. There is nothing logically contradictory in the two statements that make up his decree; nevertheless, it cannot be carried out in practice. That is how the paradox appeared to Donald John O'Connor, a philosopher at the University of Exeter, who was the first to discuss the paradox in print (*Mind*, July 1948). O'Connor's version of the paradox concerned a military commander who announced that there would be a Class A blackout during the following week. He then defined a

Class A blackout as one that the participants could not know would take place until after 6 P.M. on the day it was to occur.

"It is easy to see," wrote O'Connor, "that it follows from the announcement of this definition that the exercise cannot take place at all." That is to say, it cannot take place without violating the definition. Similar views were expressed by the authors of the next two articles (L. Jonathan Cohen in *Mind* for January 1950, and Peter Alexander in *Mind* for October 1950), and even by George Gamow and Marvin Stern when they later included the paradox (in a man-to-be-hanged form) in their book *Puzzle Math* (Viking, 1958).

Now, if this were all there was to the paradox, one could agree with O'Connor that it is "rather frivolous." But, as Scriven was the first to point out, it is by no means frivolous, and for a reason that completely escaped the first three authors. To make this clear, let us return to the man in the cell. He is convinced, by what appears to be unimpeachable logic, that he cannot be hanged without contradicting the conditions specified in his sentence. Then on Thursday morning, to his great surprise, the hangman arrives. Clearly he did not expect him. What is more surprising, the judge's decree is now seen to be perfectly correct. The sentence can be carried out exactly as stated. "I think this flavour of logic refuted by the world makes the paradox rather fascinating," writes Scriven. "The logician goes pathetically through the motions that have always worked the spell before, but somehow the monster, Reality, has missed the point and advances still."

In order to grasp more clearly the very real and profound linguistic difficulties involved here, it would be wise to restate the paradox in two other equivalent forms. By doing this we can eliminate various irrelevant factors that are often raised and that cloud the issue, such as the possibility of the judge's changing his mind, of the prisoner's dying before the hanging can take place, and so on.

The first variation of the paradox, taken from Scriven's article, can be called the paradox of the unexpected egg. Imagine that you have before you ten boxes labeled from 1 to 10. While your back is turned, a friend conceals an egg in one of the boxes. You turn around. "I want you to open these boxes one at a time," your friend tells you, "in serial order. Inside one of them I guarantee that you will find an unexpected egg. By 'unexpected' I mean that you will not be able to deduce which box it is in before you open the box and see it."

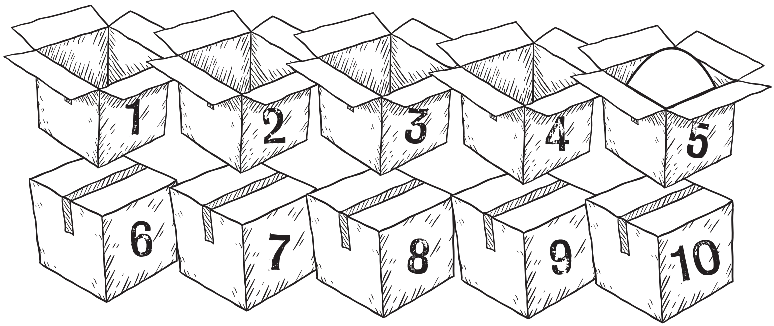

Figure 2. The paradox of the unexpected egg.

Assuming that your friend is absolutely trustworthy in all statements, can this prediction be fulfilled? Apparently not. Your friend obviously will not put the egg in box 10, because after you have found the first nine boxes empty you will be able to deduce with certainty that the egg is in the only remaining box. This would contradict your friend's statement. Box 10 is out. Now consider the situation that would arise if your friend were so foolish as to put the egg in box 9. You find the first eight boxes empty. Only 9 and 10 remain. The egg cannot be in box 10. Ergo it must be in 9. You open 9. Sure enough, there it is. Clearly it is an *expected* egg, and so your friend is again proved wrong. Box 9 is out. But now you have started on your inexorable slide into unreality. Box 8 can be ruled out by precisely the same logical argument, and similarly boxes 7, 6, 5, 4, 3, 2 and 1. Confident that all ten boxes are empty, you start to open them. What have we here in box 5? A totally unexpected egg! Your friend's prediction is fulfilled after all. Where did your reasoning go wrong?

To sharpen the paradox still more, we can consider it in a third form, one that can be called the paradox of the unexpected spade. Imagine that you are sitting at a card table opposite a friend who is holding all the 13 spades. After shuffling them, fanning them so you can't see the faces and dealing a single card face down on the table, your friend asks you to name slowly the 13 spades, starting with the ace and ending with the king. Each time you fail to name the card on the table your friend will say "No," and when you name the card correctly, "Yes."

"I'll wager a thousand dollars against a dime," your friend says, "that you will not be able to deduce the name of this card before I respond with 'Yes.'"

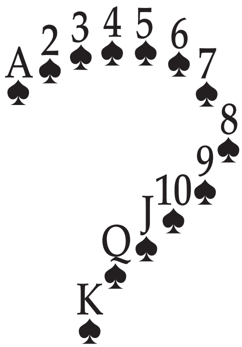

Figure 3. The paradox of the unexpected spade.

Assuming that your friend plans not to lose money, is it possible that the king of spades was placed on the table? Obviously not. After you have named the first 12 spades, only the king will remain. You will be able to deduce the card's identity with complete confidence. Can it be the queen? No, because after you have named the jack only the king and queen remain. It cannot be the king, so it must be the queen. Again, your correct deduction would win you $1,000. The same reasoning rules out all the remaining cards. Regardless of what card it is, you should be able to deduce its name in advance. The logic seems airtight. Yet it is equally obvious, as you stare at the back of the card, that you have not the foggiest notion which spade it is!

Even if the paradox is simplified by reducing it to two days, two boxes, two cards, something highly peculiar continues to trouble the situation. Suppose your friend holds only the ace and deuce of spades. It is true that you will be able to collect your bet if the card is the deuce. Once you have named the ace and it has been eliminated you will be able to say: "I deduce that it's the deuce." This deduction rests, of course, on the truth of the statement "The card before me is either the ace or the deuce of spaces." (It is assumed by everybody, in all three paradoxes, that the man *will* be hanged, that there *is* an egg in a box, that the cards *are* the cards designated.) This is as strong a deduction as mortal man can ever make about a fact of nature. You have, therefore, the strongest possible claim to the $1,000.

Suppose, however, your friend puts down the ace of spades. Cannot you deduce at the outset that the card is the ace? Surely your friend would not risk $1,000 by putting down the deuce. Therefore it *must* be the ace. You state your conviction that it is. Your friend says "Yes." Can you legitimately claim to have won the bet?

Curiously, you cannot, and here we touch on the heart of the mystery. Your previous deduction rested only on the premise that the card was either the ace or the deuce. The card is not the ace; therefore it is the deuce. But now your deduction rests on the same premise as before plus an additional one, namely on the assumption that your friend spoke truly; to say the same thing in pragmatic terms, on the assumption that your friend will do everything possible to avoid paying you $1,000. But if it is possible for you to deduce that the card is the ace, your friend will lose money just as surely as if it were the deuce. Since your friend loses it either way, there is no rational basis for picking one card rather than the other. Once you realize this, your deduction that the card is the ace takes on an extremely shaky character. It is true that you would be wise to bet that it is the ace, because it probably is, but to win the bet you have to do more than that: you have to prove that you have deduced the card with iron logic. This you cannot do.

You are, in fact, caught up in a vicious circle of contradictions. First you assume that your friend's prediction will be fulfilled. On this basis you deduce that the card on the table is the ace. But if it is the ace, the prediction is falsified. If the prediction cannot be trusted, you are left without a rational basis for deducing the name of the card. And if you cannot deduce the name of the card, the prediction will certainly be confirmed. Now you are right back where you started. The whole circle begins again. In this respect the situation is analogous to the vicious circularity involved in a famous card paradox first proposed by the English mathematician P. E. B. Jourdain in 1913 (see Figure 4). Since this sort of reasoning gets you no further than a dog gets in chasing its tail, you have no logical way of determining the name of the card on the table. Of course, you may *guess* correctly. Knowing your friend, you may decide that it is highly probable that the card is the ace. But no self-respecting logician would agree that you have "deduced" the card with anything close to the logical certitude involved when you deduced that it was the deuce.

The flimsiness of your reasoning is perhaps seen more clearly if you return to the 10 boxes. At the start you "deduce" that the egg is in box 1, but box 1 is empty. You then "deduce" it to be in box 2, but

THE SENTENCE
ON THE OTHER SIDE
OF THIS CARD
IS TRUE

THE SENTENCE
ON THE OTHER SIDE
OF THIS CARD
IS FALSE

Figure 4. P. E. B. Jourdain's card paradox.

box 2 is empty also. Then you "deduce" box 3, and so on. (It is almost as if the egg, just before you look into each box in which you are positive it must be, were cleverly transported by secret trap doors to a box with a higher number!) Finally you find the "expected" egg in box 8. Can you maintain that the egg is truly "expected" in the sense that your deduction is above reproach? Obviously you cannot, because your seven previous "deductions" were based on exactly the same line of reasoning, and each proved to be false. The plain fact is that the egg can be in any box, *including the last one*.

Even after having opened nine empty boxes, the question of whether you can "deduce" that there is an egg in the last box has no unambiguous answer. If you accept only the premise that one of the boxes contains an egg, then of course an egg in box 10 can be deduced. In that case, it is an expected egg and the assertion that it would not be is proved false. If you also assume that your friend spoke truly when he said the egg would be unexpected, then nothing can be deduced, for the first premise leads to an expected egg in box 10 and the second to an unexpected egg. Since nothing can be deduced, an egg in box 10 will be unexpected and both premises will be vindicated, but this vindication cannot come until the last box is opened and an egg is found there.

We can sum up this resolution of the paradox, in its hanging form, as follows. The judge speaks truly and the condemned man reasons falsely. The very first step in his chain of reasoning – that he cannot be hanged on the last day – is faulty. Even on the evening of the next-to-last day, as explained in the previous paragraph with reference to the egg in the last box – he has no basis for a deduction. This is the main point of Quine's 1953 paper. In Quine's closing words, the condemned man should reason: "We must distinguish four cases: first, that I shall be hanged tomorrow noon and I know it now (but I do not); second, that I shall be unhanged tomorrow noon and know it now (but I do not); third, that I shall be unhanged tomorrow noon and do not know it now; and fourth, that I shall be hanged tomorrow noon and do not know it now. The latter two alternatives are the open possibilities, and the last of all would fulfill the decree. Rather than charging the judge with self-contradiction, therefore, let me suspend judgment and hope for the best."

The Scottish mathematician Thomas H. O'Beirne, in an article with the somewhat paradoxical title "Can the Unexpected *Never* Happen?" (*The New Scientist*, May 25, 1961), has given what seems to me an excellent analysis of this paradox. As O'Beirne makes clear, the key to resolving the paradox lies in recognizing that a statement about a future event can be known to be a true prediction by one person but not known to be true by another until after the event. It is easy to think of simple examples. Someone hands you a box and says: "Open it and you will find an egg inside." *He* knows that his prediction is sound, but *you* do not know it until you open the box.

The same is true in the paradox. The judge, the man who puts the egg in the box, the friend with the 13 spades – each knows that his prediction is sound. But the prediction cannot be used to support a chain of arguments that results eventually in discrediting the prediction itself. It is this roundabout self-reference that, like the sentence on the face of Jourdain's card, tosses the monkey wrench into all attempts to prove the prediction unsound.

We can reduce the paradox to its essence by taking a cue from Scriven. Suppose a man says to his wife: "My dear, I'm going to surprise you on your birthday tomorrow by giving you a completely unexpected gift. You have no way of guessing what it is. It is that gold bracelet you saw last week in Tiffany's window."

What is the poor wife to make of this? She knows her husband to be truthful. He always keeps his promises. But if he does give her the gold bracelet, it will not be a surprise. This would falsify his prediction. And if his prediction is unsound, what *can* she deduce? Perhaps he will keep his word about giving her the bracelet but violate his word that the gift will be unexpected. On the other hand, he may keep his word about the surprise but violate it about the bracelet and give her instead, say, a new vacuum cleaner. Because of the self-refuting character of her husband's statement, she has no rational basis for choosing between these alternatives; therefore she has no rational basis for expecting the gold bracelet. It is easy to guess what happens. On her birthday she is surprised to receive a logically unexpected bracelet.

He knew all along that he could and would keep his word. *She* could not know this until after the event. A statement that yesterday appeared to be nonsense, that plunged her into an endless whirlpool of logical contradictions, has today suddenly been made perfectly true and noncontradictory by the appearance of the gold bracelet. Here in the starkest possible form is the queer verbal magic that gives to all the paradoxes we have discussed their bewildering, head-splitting charm.

ADDENDUM

A great many trenchant and sometimes bewildering letters were received from readers offering their views on how the paradox of the unexpected hanging could be resolved. Several went on to expand their views in articles that are listed in the bibliography for this chapter. (Ordinarily I give only a few select references for each chapter, but in this case it seemed that many readers would welcome as complete a listing as possible.)

Lennart Ekbom, who teaches mathematics at Östermalms College, in Stockholm, pinned down what may be the origin of the paradox. In 1943 or 1944, he wrote, the Swedish Broadcasting Company announced that a civil-defense exercise would be held the following week, and to test the efficiency of civil-defense units, no one would be able to predict, even on the morning of the day of the exercise, when it would take place. Ekbom realized that this involved a logical paradox, which he discussed with some students of mathematics and philosophy at Stockholm University. In 1947 one of these students visited Princeton, where he heard Kurt Gödel, the famous mathematician, mention a variant of the paradox. Ekbom adds that he originally believed the paradox to be older than the Swedish civil-defense announcement, but in view of Quine's statement that he first heard of the paradox in the early forties, perhaps this was its origin.

The following two letters do not attempt to explain the paradox, but offer amusing (and confusing) sidelights. Both were printed in *Scientific American's* letters department, May 1963.

SIRS:

In Martin Gardner's article about the paradox of the unexpected egg he seems to have logically proved the impossibility of the egg being in any of the boxes, only to be amazed by the appearance of the egg in box 5. At first glance this truly is amazing, but on thorough analysis it can be proved that the egg will always be in box 5.

The proof is as follows:

Let S be the set of all statements.

Let T be the set of all true statements.

Every element of S (every statement) is either in the set T or in the set $C = S - T$, which is the complement of T, and not in both.

Consider:

(1) Every statement within this rectangle is an element of C.
(2) The egg will always be in box 5.

Statement (1) is either in T or in C and not in both.

If (1) is in T, then it is true. But if (1) is true, it asserts correctly that every statement in the rectangle, including (1), is in C. Thus, the assumption that (1) is in T implies that (1) is in C.

Contradiction

If (1) is in C, we must consider two cases: the case that statement (2) is in C and the case that (2) is in T.

If (2) is in C, then both (1) and (2), that is, every statement in the rectangle, is an element of C. This is exactly what (1) asserts, and so