Estratto da

M. Barra e A. Zanardo (a cura di), *Atti degli incontri di logica matematica* Volume 5, Roma 6-9 aprile 1988.

Disponibile in rete su http://www.ailalogica.it

LOGIC PAZZLES

RAYMOND M. SMULLYAN
New York USA

Logic puzzles constitute one introduction to mathematical logic that has proved attractive to many a student. We will consider several such puzzles here (and the relation of some of them to the teaching of logic) and we shall conclude with a new paradox.

Let us start with a logic puzzle that has a startling conclusion: Suppose two people, A and B each made you an offer; the problem is which of the two offers you would prefer to accept. A says: "You are to make a statement. If the statement is true, I will give you exactly ten dollars. If the statement is false, then I will give you either more or less than ten, but not exactly ten." Now comes B's offer: "You are to make a statement. Regardless of whether the statement is true or false, I will give you more than ten dollars." Which of the two offers would you prefer to accept?

Most people prefer B's offer on the grounds that with this offer you are guaranteed more than ten dollars, whereas with A's offer, the most you could be <u>sure</u> of is ten. I, however, prefer A's offer, for if A makes me his offer, I would say: "You will give me neither exactly ten dollars, nor exactly a million dollars." Now, my statement is either true or false.

If it is true, then A must give me exactly ten dollars as agreed, but his doing so would make the statement false! And so it is contradictory to assume the statement true. Therefore the statement must be false. Since it is false that A will give me neither ten nor a million, then A must give me either ten or a million. But it is against the rules that he gives me (exactly) ten dollars for a false statement, and so he owes me a million dollars! [This puzzle is taken from [5]]

The above puzzle is related to the following one (which in turn is related to Gödel's incompleteness theorem). On the first day of my introductory logic class I place a penny and a quarter on the table and say to the student: "You are to make a statement. If the statement is true, then I will give you either the penny or the quarter (but not both). If your statement is false, then I won't give you either coin." The problem is: What statement can the student make that will force me to give him the quarter?

One solution is that the student says: "You will not give me the penny." If I should give him the penny, then I would falsify his statement, hence I would be giving him a coin for a false statement, which is against the rules. Therefore I cannot give him the penny, hence his statement is true. But then I must give him one of the two coins for having made a true statement, and since I cannot give him the penny, I must give him the quarter.

If we think of the penny as corresponding to a sentence that is provable in a given mathematical system and the quarter as corresponding to a sentence that is true but not provable in the system, then the sentence "You will not give me the penny" is the analogue of Gödel's famous sentence that asserts its own non-

provability in the system. Another puzzle that illustrates Gödel's principle even better is the following.

We visit the Island of Knights and Knaves (of the type considered in [1] - [5]) in which each inhabitant is either a knight or a knave, but not both. Knights make only true statements and knaves make only false ones. On this island, no inhabitant can claim that he is not a knight (for a knight would never falsely deny that he is a knight, and a knave would never make the true statement that he is not a knight). Now suppose a logician visits the island and is told by a native: "You can never prove that I am a knight." Let us assume that the logician is completely accurate in his beliefs; everything he can prove is really true. Is the inhabitant a knight or a knave?

Well, if the logician could prove that the native is a knight, then the native's statement would be false, hence the native would really be a knave, hence the logician would have falsely proved that the native is a knight, which is contrary to the given condition that the logician is accurate and proves only true statements. Therefore the logician can never prove that the native is a knight. Well, the native said just that, hence he must be a knight. And so the conclusion is that the native is a knight, but the logician can never prove that he is!

Suppose now that I give you the additional information that the logician knows logic as well as you and I. Now, you and I have just proved that the native is a knight; what is to prevent the logician from going through the same argument and proving that the native is a knight? But if he could prove that the native is a knight, this would automatically falsify the natives claim, thus making the native in reality a knave! So don't we get a paradox?

No, we do not! There is something we know that the logician doesn't know. Can you see what that is?

The answer is that I told you that the logician is always accurate, but I never told you that the logician knew (or could prove) that he is! Indeed, if he could prove that he is always accurate, then he could prove that the native is a knight (by the same argument we have used), and in so doing would lose its accuracy (since the native would then be a knave). And so the upshot is that under the given conditions, if the logician is wholly accurate, then he can never know (or prove) that he is. This bears a resemblance to Gödel's second theorem -- that a mathematical system of sufficient strength can never prove its own consistency, unless the system is inconsistent. [Cf. [5] for a more in-depth treatment of this point.]

We next turn to a puzzle of a different nature: Suppose a native of the island says: "This is not the first time I have said what I am now saying." Is the native a knight or a knave?"

A moment's thought will reveal that if the native is a knight, then he really has made the statement before (as he claimed), and when he made it before, he was a knight then, hence he must have made it a time before that, hence a time before that -- and so on, ad infinitum. So unless he has lived infinitely far back in the past, he must be a knave. [Alternatively, and perhaps this is simpler, since he made the statement once, there must have been a first time that he made it. But the first time he made it, it was clearly false.]

There is also a "forward" version of this problem -- a recipe for immortality. Do you wish to know how you can live

forever? It's really very simple: all you need to do are two things: First, you must make only truthful statements from now on; never say anything false. [That's a small price to pay for immortality, isn't it?] The second thing is that you say: "I'll repeat this statement tomorrow." If you do these two things, then I'll guarantee that you will live forever.

I use the above two puzzles to introduce the subject of mathematical induction. Another device I have found useful is the following: Let us imagine that we are all immortal and that we live in the good old days when the milkman would deliver milk to our door and the housewife would leave a note in an empty bottle telling the milkman what to do. Suppose the housewife leaves the following note: "If ever you leave milk on one day, be sure to leave it the next day as well." [Or alternatively: "Never leave milk one day and fail to leave it the next day."] Well, the milkman could go for a thousand years without leaving milk and he would not have violated the housewife's order. But if one day on a whim he should decide to leave milk, then he is clearly committed evermore. And so the housewife's order was not good enough; what she should have written was two things:
(1) If ever you leave milk one day, leave it the next day as well;
(2) leave milk today. These two things would guarantee permanent delivery.

A friend of mine (Alan Tritter, a computer scientist) thought of the following cute variant of this (which illustrates what might be called the Turing Machine, or the Recursive approach). His note reads: "Leave milk today and read this note again tomorrow."

This brings us to the subject of infinity. This subject could (and <u>should</u>) be taught in secondary schools (if not earlier!). Beginners -- even those who claim to be "poor in math" -- are usually completely fascinated by this topic. After going through the usual introductory bit about Hilbert's Hotel, the students usually ask me to explain more clearly what is meant by a set being infinite. Well, I first explain what is meant by a 1-1 correspondence. [As an example, I point out that if one sees that every seat of a theatre is taken, and that no one is standing, then without having to count either the number of seats, or the number of people, one knows that the two numbers are the same.] I of course presuppose that the students are familiar with the natural numbers 0, 1, 2,..., n,... . I then explain that for any positive natural number n, a set of objects is said to have n elements if it can be put into a 1-1 correspondence with the set of natural numbers from 1 to n (and that establishing the correspondence is what is commonly called <u>counting</u>), and a set is said to have zero elements if it has no elements at all. I then define a set to be <u>finite</u> if there is a natural number n such that the set has n elements; otherwise the set is called <u>infinite</u>. [Thus definition seems more intuitive to beginners than the definition of a set being infinite if it can be put into a 1-1 correspondence with a proper subset of itself.] I then ask the students to make an intuitive guess as to the answer to the following question: Can any two infinite sets be put into a 1-1 correspondence with each other, or do infinite sets come in different sizes? [Most students, so far, have guessed that any two infinite sets <u>are</u> of the same size!] I then explain that Georg Cantor (the true father of the theory of infinite sets) addressed himself to this problem and at

first believed that any two infinite sets were of the same size, but later realized (and proved) that infinite sets <u>do</u> come in different sizes. What Cantor first did was to consider sets that on the surface <u>appeared</u> to be uncountable (larger than the set of positive integers) but which turned out to be countable after all. I illustrate this in the following manner (which has a nice constructive flavor).

I ask the students to imagine that they are in the underworld and that I am the devil. I write down a positive integer on a piece of paper and say: "Each day you have one and only one guess as to what number I have written. If and when you guess it correctly, you go free." What strategy will guarantee eventual freedom? Of course all the students solve that: On the first day you ask if the number is 1; on the second day, if the number is 2, and so forth. Eventually you must come to my number. The second test is but a shade more difficult: This time the devil writes down either a positive integer or a negative integer and the victim must guess what is written (one guess per day). Must of the students solve that easily enough (they count the integers in the order 1, -1, 2, -1,.., 3, -3,...)

The next test is more interesting. The devil now says: "This time I have written down <u>two</u>positive integers (maybe the same one repeated). Each day you have one and only one guess as to what they <u>both</u> are. It's no good if you guess one of them one day and another on another day; you must guess both of them on the same day." <u>Now</u> is there a strategy which will surely set you free eventually? At this point many of the students are doubtful. They reason that there are infinitely many possibilities for one of the numbers and with each such possibility, there are infinitely many possibilities for the other, and so the case seems hopeless. I then tell them that there is such a strategy and I leave it as a homework problem to try and figure one out. The next session, the brighter students have solved it: They realize that there is only one pair in which the highest number is 1 (namely (1, 1)); two pairs in which the highest is 2 (namely (1, 2) and 2, 2)), and so forth. I then point out that if the victim had to guess the order in which the numbers were written, as well as the numbers themselves, it might take about twice as long to get out, but still he could be sure of getting out. And that is how Cantor enumerated the rationals (the integral numerator and the integral denominator).

For the next test, the devil writes down a finite set of positive integers. He doesn't tell you how many numbers are in the set, nor what the highest number of the set is. Now is there a strategy for getting out? Only the very brightest students get this one. [For every positive n, there are only finitely many sets whose highest number is n -- namely, 2^{n-1}.]

And so we see how one can enumerate the set of all finite sets of positive integers. Then I explain that if one wishes to enumerate the set of all sets of positive integers, it is simply impossible to do so. And I illustrate Cantor's proof of this as follows: Imagine that one has a book with infinitely many pages -- page 1, page 2, .., page n, . . . On each page is described a set (finite or infinite) of positive integers. If every set of positive integers is listed somewhere in the book, then the owner gets a grand prize. But if the owner brings me the book(1), then

(1) Actually, the book might be a bit heavy (unless the pages get progressively thinner and thinner)!

without even opening it, I can name a set that is not listed anywhere in the book -- namely, the set of all n such that n does not belong to the set listed on page n. This, of course, is Cantor's proof.

Now for a little paradox: Suppose the very description: "The set of all n such that n does not belong to the set described on page n" -- suppose this description appears on one of the pages -- say, page 13. Does 13 belong to this set or not? Either way, we get a clear contradiction. How does one get out of it?

The solution is quite simple: If that phrase should actually appear on any page of the book, it wouldn't describe any set at all;it would be what is called a pseudo-description (for if it were a genuine description, then we would have a contradiction!)

Now, let me conclude with a more baffling paradox: Suppose we have another book, and on each page we are allowed to have either a genuine description or a pseudo-description of a set. [Some of the descriptions may be genuine and others not.] Now consider the following description: "The set of all n such that the description on page n is genuine and n does not belong to the set described on page n."

Is the above description genuine or not? Let me prove to you that it is. To prove this, it suffices to show that for every n, we have a definite rule for determining whether n belongs to our set or not. Well, take any n. Either the description on page n is genuine or it isn't. If it isn't, then n automatically does not belong to our set. Now suppose the description on page n is genuine. Then it names some definite set, and n then belongs to our set if and only if it doesn't belong to that definite set. And so the description *is* genuine. Now, suppose that that genuine description occurs somewhere in the book -- say on page 13. I've already proved that the description is genuine, and so 13 belongs to the set described on page 13 if and only if it doesn't belong to the set described on page 13. How does one get out of this one?

References

[1] - [5] below are all by the present author.

[1] *What Is The Name Of This Book?*
Prentice Hall, 1978

[2] *The Lady Or The Tiger?*
Alfred Knopf, 1982

[3] *Alice In Puzzle-Land*
William Morrow &Co., 1982

[4] *To Mock a Mockingbird*
Alfred Knopf, 1985

[5] *Forever Undecided -- A Puzzle Guide to Gödel*
Alfred Knopf, 1987

Solution To The Second Paradox

The solution is that the very notion of a genuine description is not well defined! One can define the notion of a genuine description only relative to a given language.

The situation is analogous to the fact that *truth* is definable only relative to a given language (and in many cases, not definable in the same language, as was demonstrated by Alfred Tarski).