

The Unexpected Hanging Paradox

Aditya Gupta, R K Shishir

Introduction

The unexpected hanging paradox is a logical puzzle which was first stated by Martin Gardner in 1963 in the Scientific American magazine. This paradox has been an active topic of debate for decades and has been discussed by many philosophers and mathematicians. Let us see what this paradox is all about and try to understand the various arguments that have been put forward to explain it.

The Paradox statement

We have a prisoner who is currently on death row for his crimes. On a Sunday, he is presented in the court of a judge who sentences him to be hanged at 12 noon on a weekday in the coming week. However, he also adds that the prisoner's execution would be a surprise to the prisoner.

We assume here that the judge always speaks the truth. We also assume that 'surprise' refers to the prisoner not expecting to be hanged on that day. The judge is not allowed to surprise the prisoner by doing something else, like changing the method of execution.

Upon hearing this, the prisoner is left hanging (pun intended). However, the prisoner is also a mathematician and is thus determined to find a way to escape death. He goes back to his cell on Sunday and starts analysing his situation. Soon, he comes up with an absurd conclusion.

The prisoner's expectation

After thinking about the judge's statement for a while, the prisoner concludes that he cannot be hanged at all. He believes that there is no day in the coming week when the hanging could take place without him expecting it. Let's see why he thinks so.

Initial argument

He starts counting the days of week backwards, starting from Friday. Suppose he was to be executed on Friday. For that to happen, he would have to be alive on the

Thursday night. But if he was alive on Thursday night, the only possible day left for his execution would be Friday. He would be able to deduce on Thursday night that he will be hanged the next day. Now, an execution on Friday no longer remains a surprise to the prisoner. Since the judge promised that the execution would be a surprise, it cannot happen on Friday.

Taking it forward

Having eliminated one day, the prisoner now tries to come up with similar explanations for the other days of the week too. Right now, the only possible days for his execution are Monday to Thursday. But once again, if he is alive on Wednesday night, then the only possible day he can be executed on is Thursday. Thus, his execution on Thursday also does not remain a surprise to him. Therefore, he can not be executed on Thursday either.

The final conclusion

The prisoner continues this argument for the remaining days of the week and concludes that he cannot be executed on any day of the week. He believes that the judge's statement has made it impossible for him to be hanged. He is now confident that he will be alive at the end of the week.

However, things do not go as he planned. On Wednesday afternoon, the executioner knocks on the prisoner's door and he is hanged. The prisoner is shocked. He was so confident that he would be alive at the end of the week. How did he go wrong? Why does this paradoxical situation arise? Let us try to understand this.

Resolving the paradox

There are a few arguments that have been put forward to explain this paradox. However, this paradox is still a topic of debate and there is no consensus on which argument is correct. Let us look at some of the arguments.

A naïve resolution

Let us first look at a very obvious (and almost laughable) way to see this problem. After his analysis, the prisoner came to the conclusion that he can not be hanged on any day. This means that on any given weekday, the prisoner is not expecting to be hanged. This makes his hanging on any weekday a surprise for him. Thus, the judge's statement is satisfied and the prisoner is hanged. By overanalysing the situation, the prisoner doomed himself to destined death.

The judge's statement

Another way to think about this is by reevaluating what the judge said. The judge said that the prisoner's execution would be a surprise to him. However, from his statement, it is not clear as to 'when' the prisoner would be surprised. The prisoner assumed that he would be surprised on the day of his execution. However, by surprise, the judge could be implying that the prisoner will not be able to deduce the day of his execution on Sunday night. The reasoning used by the prisoner needs some extra information, such as him surviving till a particular day. However, he does not have any of this information on Sunday night. Since the assumptions he makes can not be made on Sunday night, any further arguments derived from them are also invalid. Hence, all of the prisoner's arguments are invalid.

To give an example, the executioner could secretly roll a 5 sided dice on Sunday night and use it to determine the day on which he would execute the prisoner. In this case, since the prisoner does not know the outcome of the dice roll, he has no way of predicting the day of his execution.

This problem mainly arises due to the gap between the english language and formal mathematical language. The statement made by the judge is not precise enough to be interpreted in a mathematical way. This prevents the prisoner from making logical deductions based on the judge's statement.

Self reference

As we saw above, the meaning of surprise is not very clear mathematically. We can try to define surprise more mathematically as "not being able to deduce using the given statements". Using this definition, the judge's statement now becomes "The prisoner will be hanged next week and the day of the hanging can not be deduced by using this statement".

This might seem like a good way to define the judge's statement but this definition actually leads to a serious issue called self reference. Let us see how.

A self referential statement is a statement that refers to its own attributes or characteristics. For example, the statement "This statement is false" (commonly known as the Liar's paradox) is self referential. If the statement is true, then it is false. If it is false, then it is true. Thus, this statement is neither true nor false. Self referential statements like these generally lead to paradoxical situations.

The prisoner assumed the judge's statement to be true and tried to deduce the day of his execution. Using seemingly correct logical methods, the prisoner was able to conclude that he won't be hanged. However, by proving that he won't be hanged, the prisoner has managed to prove the judge's statement to be false. This is a paradoxical situation. The prisoner has managed to prove the judge's statement to be false, even though he assumed it to be true. This is similar to the Liar's paradox.

Conclusion

Attempting to write the judge's statement in a mathematical way has led to a self referential statement. The statement has no definitive truth value due to its self referential nature. Thus, the prisoner's arguments are invalid. This is an inherent flaw in the structure of mathematics and logic. Self referential statements like these tend to break the rules of logic and mathematics. This is why this paradox is still a topic of debate and there is no consensus on which argument is correct.

What if the time of the execution was not specified?

An interesting variation that we can bring to this paradox is by removing the time of execution from the judge's statement. Suppose the judge had said that the prisoner would be hanged on a weekday in the coming week and that the day of the hanging would be a surprise to the prisoner. Here, the judge does not specify the time of the execution, meaning the prisoner could be hanged at any moment of any given day.

This prevents the prisoner from starting his inductive argument. His execution on Friday can still be a surprise since at no given time, the prisoner can know if he could be hanged the very next moment. A moment of time is an infinitesimally small quantity. Since there are an infinite number of real numbers between 0 and 1, between 12:00 pm and 12:01 pm, there are an infinite number of real times where the execution can take place. These infinite possibilities make it impossible for the prisoner to be sure about his execution at any given moment.