

A naive research of the Surprise Test Paradox

YANG Zonghan

April 4, 2020

1 Introduction

A teacher announces that there will be a surprise test next week. A student objects that this is impossible: 'The class meets on every weekday. If the test is given on Friday, then on Thursday I would be able to predict that the test is on Friday – It would not be a surprise. Can the test be given on Thursday? No, because on Wednesday I would know that the test will not be on Friday – thanks to the previous reasoning – and know that the test was not on Wednesday – thanks to memory. Therefore, on Wednesday I could foresee that the test will be on Thursday. A test on Thursday would not be a surprise. Similarly, a Wednesday, Tuesday, Monday test would also fail to be a surprise. Therefore, it is impossible for there to be a surprise test.'

Can the teacher fulfill her announcement? Say, is that argument student invoked incorrect?

In some versions of the story, the test was decided to take place on Tuesday, becoming a real surprise to students. However, if we think more carefully, the test will be a surprise to the student no matter when it took place, because student believed in the impossibility. This is, obviously, a paradox. In fact it's one of the most famous paradoxes, known as Unexpected Hanging Paradox / Surprise Test Paradox / The Tiger Paradox, one kind of epistemic¹ paradoxes.

A correct deduction ought to lead to a correct answer. So there are two possibilities:

1. The teacher lied.
2. The student's deduction is wrong.

So mainly we want to investigate the problem, and find out the 'invisible' fault in it.

2 History of similar problems

The discussion of such problems can be traced back back a long time ago, which was not specifically mentioned who raise this question first. One is believed to be the first printed form of this argument is published on the philosophy journal *Mind*. It was used as an introduction to discuss many pragmatic paradoxes².

The forms of the legend are varied, too. In the above paper it's about 'Class A blackout' exercise happened in the military. The most famous one is that judge tells a condemned prisoner that he will be hanged 'surprisingly' someday in the next week. Other versions of the paradox replace the death sentence with a surprise fire drill, examination, pop quiz, A/B test launch or a lion behind a door. [2] In most Chinese versions of that story, it's called The Tiger Paradox, whose story is about an unexpected tiger behind the cage.

¹episteme is Greek for knowledge

²pragmatic means practical, and pragmatic paradox is a statement that is falsified by its own utterance

PRAGMATIC PARADOXES

PHILOSOPHERS have spent a good deal of time and trouble in elucidating the so-called "logical paradoxes". And although their efforts have not yet been completely successful, these paradoxes are now a good deal less puzzling than they were when they were first propounded. But there is another class of paradoxes which has received less attention, partly no doubt because they do not appear at first sight to raise any interesting technical questions of logic or to point the way to new technical developments. Nevertheless, these "pragmatic paradoxes" as they have been called, are worth examination, although I shall not do any more here than draw attention to some of their characteristics and commend them to the attention of philosophers.

Consider the following case. The military commander of a certain camp announces on a Saturday evening that during the following week there will be a "Class A blackout". The date and time of the exercise are not prescribed because a "Class A blackout" is defined in the announcement as an exercise which the participants cannot know is going to take place prior to 6.0 p.m. on the evening in which it occurs. It is easy to see that it follows from the announcement of this definition that the exercise cannot take place at all. It cannot take place on Saturday because if it has not occurred on one of the first six days of the week it must occur on the last. And the fact that the participants can know this violates the condition which defines it. Similarly, because it cannot take place on Saturday, it cannot take place on Friday either, because when Saturday is eliminated Friday is the last available day and is, therefore, invalidated for the same reason as Saturday. And by similar arguments, Thursday, Wednesday, etc., back to Sunday are eliminated in turn, so that the exercise cannot take place at all.

Now though there is an obvious fault of definition in this case, the fault is not a fault of logic in the sense that the definition is formally self-contradictory. It is merely *pragmatically* self-refuting. The conditions of the action are defined in such a way that their publication entails that the action can never be carried out. Now why should philosophers be interested in this sort of situation? It seems to me that there are a number of examples of such paradoxes which can arise in philosophical discussions and which deserve the attention of philosophers even if the rather frivolous example I have just given does not interest them.

If I say "I do not exist", my statement is L-false as the word "I" in a given context functions as a proper name. But suppose I say "I remember nothing at all". This is not logically self-contradictory. And it is not merely a false factual statement like "X remembers nothing" where X is giving accurate answers to questions about his past life. But it is like an L-false statement in

358

PRAGMATIC PARADOXES by D. J. O'CONNOR, 1948, [3]

3 A game theory's perspective

Suppose now we have a game. There are $n > 1$ choices under some data and laws. An agent, 'Predictor', who has all the data, laws, and calculating capacity needed to predict the choices of others. The agent, 'Avoider', whose dominant motivation is to make a choice and avoid prediction, has access to the same data, laws, and calculating capacity as Predictor. Predictor must make a prediction and conceal his prediction. Then Avoider should make a choice to avoid that happen. In the original problem, teacher is the Avoider and student is the Predictor. It was the model Michael Scriven used in his article [1].

Why we need the above model which seems trivial? Actually we want to show a fact that we need.

Fact 3.1. *There's no good deterministic policy for Predictor.*

Proof. If such policy exists, then Avoider simply copy that policy and conclude a result. All Avoider need to do is to choose any other choices and win the game. \square

Applying that argument to our specific problem, what I come up with is like *Lasagna*³.



Lasagna, Wikipedia [5]

Making prediction and avoiding prediction is a kind of climbing on the thinking layers, like "I know it can't be Friday, you know I know it can't be Friday, I know you know I know it can't be Friday..."

³A multi-layered cake(千层饼), especially in Italy. Similar snacks varied in different places.

Predictor is not able to stay at any *determined* layer. Once Predictor made such decision, Avoider can go simply *one layer higher* – think just further than Predictor, to defeat such policy. What's worse, Predictor can only defeat those are *one layer lower* – think just behind it, so the Avoider can be either too clever, or even too dumb to defeat.

Thus, Predictor shall use a random policy to predict. Nevertheless, with randomness it's impossible for Predictor itself to announces that it will **absolutely** win, since [s]he will faithfully toss a coin, roll a dice or something like that to make prediction which ought not to be foresee. Therefore, we have:

Fact 3.2. *There's no policy for Predictor to absolutely win the game.*

So, if we say 'know' is about the certainty of the test date, student can't *know* the date at all if [s]he is as clever as teacher. In such situation, student *learned nothing* about the date of the test in the announcement and thus, a surprise.

But from this argument, the question is not answer yet: how can we explain the 'No Friday Argument'? Is it correct or not?

4 The 'No Friday Argument'

In this part we'd like to show the 'No Friday Argument' is wrong due to something missing. I am new to philosophy, there are only naive ideas that I concluded from article I found but not formally stated propositions.

Let's reconsider the No Friday Argument.

No Friday Argument

1. There will be a surprise exam during the weekdays.
2. A test on Friday will not be a surprise if student expects the test on Thursday night.
3. From 1, 2, no test will be given on Friday.

Well, it looks good. So where's the problem? There are some arguments.

First, let's just consider Thursday night. It's the time that we can equivalently state the first statement as

There will be a surprise test *tomorrow*.

How can it make sence? 'There will be a test tomorrow' is contradicted to 'The test is surprising'. This inconsistency suggests that it is irrational for the student to accept the No Friday Argument's tacit assumption about the teacher's announcement.

Second, if the No Friday Argument is correct that an exam on Friday will not be a surprise, then the student is no longer justified in believing on Thursday night the teacher's announcement, in which case an exam on Friday *will* be a surprise – a contradiction to the No Friday Argument.

Thus, the whole argument of Surprise Test Paradox is wrong, it's basically we went too far even when we've just started to argue. The No Friday Argument itself, is wrong.

5 How about self-defeating prophecies?

It seems that the teacher is dishonest, since we proved that the announcement itself is a paradox – at least on Thursday night. Also, the first published reaction to the surprise test paradox was to endorse the student's elimination argument and criticize on the teacher's behavior. D. J. O'Connor regarded the teacher's announcement as self-defeating. "If the teacher had not announced that there would be a surprise test, the teacher would have been able to give the surprise test. The pedagogical moral of the paradox would then be that if you want to give a surprise test do not announce your intention to your students!" [3]

But the thing is not as simple as Mr. O'Connor thought. True, the teacher's announcement does compromise one aspect of the surprise: Students now know that there will be a test. But this compromise is not itself enough to make the announcement self-falsifying. The existence of a surprise test has been revealed but perhaps that allows surviving uncertainty as to which day the test will occur. The announcement of an incoming surprise aims at changing 'uninformed ignorance' into 'action-guiding awareness of ignorance' [4]. For example, a student who misses the announcement does not realize that there is a test – if no one tell them about the surprise test, the student with simple ignorance will be less prepared than classmates who know they do not know the day of the test.

Competition between accuracy and helpfulness makes it possible for an announcement to be self-fulfilling *by being self-defeating*, just like Surprise Test Paradox. One similar example is following [4].

Consider a weatherman who warns 'The midnight tsunami will cause fatalities along the shore'. Because of the warning, spectacle-seekers make a special trip to witness the wave. Some drown.

His announcement was intended to warn people. If all the people received the message and protect themselves, there should not be fatalities and the prophecy would fail. But it doesn't fail at all since some regard the warning in unusual ways. The weatherman's announcement succeeds as a prediction by backfiring as a warning.

6 Conclusion

So here comes the conclusion. A surprise exam *may* be given on any day of the testing period, including the very last day. The teacher could be somehow dishonest, since his/her words can be paradox in specific situations. But at least one thing [s]he is correct – students know little about the date the test should take place.

References

- [1] R. Buck, D. E. Harrah, F. Mosteller, and M. Scriven. Michael scriven an essential unpredictability in human behavior *. 2006.
- [2] T. Y. Chow. The surprise examination or unexpected hanging paradox. *The American Mathematical Monthly*, 105(1):41-51, 1998.
- [3] D. J. O'CONNOR. PRAGMATIC PARADOXES. *Mind*, LVII(227):358-359, 07 1948.
- [4] R. Sorensen. Epistemic paradoxes. In E. N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, summer 2018 edition, 2018.

- [5] Wikipedia contributors. Lasagne — Wikipedia, the free encyclopedia, 2020. [Online; accessed 3-April-2020].