Dean&Francis

ISSN 2959-6149

# A New Response to the Surprise Test Paradox

**Yintong Chen***

Shanghai Starriver Bilingual School, Shanghai, China

*Corresponding author: 1801010101@stu.hrbust.edu.cn

**Abstract:**

The Surprise Test Paradox is a renowned epistemic paradox that has puzzled many philosophers for a very long time. A teacher tells the students that there will be a surprise test in the following week, and the students will not know beforehand when the test will occur. However, this quickly leads to a contradiction. First, the test cannot occur on Friday, because if he hasn't been hanged by Thursday, there is only one day left, so it won't be a surprise. The test also cannot be on Thursday, since Friday is eliminated and if the test hasn't taken place by Wednesday, it must take place on Thursday, making the test unsurprising. By similar reasoning, the students could conclude that the test cannot take place on Wednesday, Tuesday, or Monday. Thus, the test cannot happen at all. Yet, the teacher still gives a surprise test on Wednesday, which (despite all of the above) is a total surprise to the students. In this article, I intend to propose a new solution to this paradox and examine the definition of "surprise", as well as discuss whether or not the student actually believes in the announcement.

**Keywords:** Surprise Test Paradox: Epistemology; Logic.

## 1. Introduction

One very classical category of paradoxes is the epistemic paradoxes. Epistemology is defined as the study of the "nature, origin, and limits of human knowledge" [1]. Similarly, epistemic paradoxes aim to explore the limits of the human mind and knowledge, and these paradoxes often poses a direct inconsistency in their logic that causes the existence of contradictions [2]. There are currently several main types of epistemic paradoxes that are still under intense discussion by philosophers; they are, respectively, the lottery, preface, knowability, and surprise examination paradox [3]. This article is primarily concerned with the last one, the surprise exam paradox, which is known as well by other names such as the unexpected hanging paradox. In this article, I aim to propose a new response to this paradox and explain why it substantially differs from many of the currently-existing responses. I aim to examine the definition of "surprising" in this paradox and show that it is in fact incomplete. I will also aim to show that many of the responses proposed by previous philosophers can be shown to be logically incomplete or inconsistent. Nevertheless, the paradox still lives on as a topic of intense debate among philosophers.

## 2. Exegesis

In this part, I attempt to introduce the surprise-exam

ISSN 2959-6149

paradox (SEP). I also aim to set up necessary notions related to this paradox that aids me in explaining my response to SEP.

Suppose that a teacher announces on Friday that a surprise test will occur on one unique day in the following week, from Monday to Friday. The teacher also claims that due to the surprising quality of the exam, the students will not be able to predict on which day the exam will occur [2]. However, this easily poses a contradiction, as the students could realize that there cannot be any test. A student could reason that the test could not be on Friday, since otherwise on Thursday evening the student could predict that the test must be on Friday (as it did not occur in the other four days) [2]. Similarly, the test cannot be on Thursday, as otherwise on Wednesday evening the student could predict that: 1) the test will not be on Friday (because of the previous reasoning), and 2) the test is not on the previous three days. This leads to the conclusion that the test must be on Thursday [2]. A similar line of reasoning could be used to prove that the test cannot be on Wednesday, Tuesday, and Monday. However, the teacher then gives the test on Wednesday, which surprises the students [2].

To be more specific, "surprising" means that on any day, the student cannot deduce for certain that the test will happen on the next day based on his memory (whether the test occurred in previous days) and the information in the announcement. In other words, the student cannot know in advance when the test will occur [2].

There are many other forms of this paradox, but perhaps the most interesting one is the version introduced by Roy Sorenson. A teacher tells five students (named A, B, C, D, and E) to line up facing one direction, with A in the front and E in the back [4]. Every student could see the backs of the students in front of them (for example, D could see the backs of A, B, C, but not E). The teacher then sticks one sticker on the back of every student, including four silver stickers and one gold sticker. The teacher then claims that the student with the gold star cannot know it unless he rips off the sticker from his back [4].

However, this also leads to a contradiction. The last student couldn't have the gold sticker, because otherwise he would be able to see that all of the other students have silver stickers, and therefore know that he has the gold sticker. A similar line of reasoning tells us that the sticker could not be on the backs of the other four students, which leads to a contradiction [4].

This is in fact a form of SEP, which becomes evident when we name the students A, B, C, D, and E with Monday, Tuesday, Wednesday, Thursday, and Friday, and the gold star be the test (though this form of the paradox takes the form of "information accumulating along a spatial axis rather than along a temporal axis") [4].

Currently, the proposed solutions to this paradox could be divided into two categories: 1) Claiming that the teacher's announcement is simply self-contradictory and that the teacher cannot fulfill his promise, and 2) Claiming that the student's line of reasoning is logically flawed. I am going to refer to these two categories by View 1 and View 2. Supporters of View 1 include philosopher such as D. J. O'Connor, who provided one of the first published responses to SEP by claiming that the teacher's announcement is "self-defeating" [2]. Supporters of View 2, including Paul Weiss, argues that the student has false assumptions or false logic [2]. This article primarily addresses the flaws present in View 1 (touching on how View 2 could successfully flip the case for view 1), while it also incorporates broader perspectives that solve SEP. touch on how View 2 could successfully flip the case for view 1. Specifically, I will discuss why the definition of "surprising" in this context should take another form which adheres more to our daily lives.

## 3. Argument

### 3.1 Part One

Let us first consider this paradox through View 1. The announcement can be divided into two statements: one, that there is a test in the following week; and two, that the test must be surprising. View 1 claims that the announcement is false. In other words, View 1 claims that either there is no test, or there is an unsurprising (expected) test that the students expect. However, in this example, both conditions do not hold. The teacher does indeed give a test in the following week, negating the former condition. The teacher also surprises the students by giving them a test that they do not expect, which is clearly seen when the students adopt the line of reasoning that is explained in the paradox. As a result, the announcement is true and not self-contradictory. Thus, I deduce that when examined in this way, View 1 does not seem to stand.

In the paradox, because of the student adopting the reasoning explained in the paradox, the student does not believe that there is a test in the following week (actually, the students may simply stay confused about whether the test will occur; however, the students will realize that the reasoning in the paradox is robust and supports the conclusion that there could not be any test. This article will also explain in Section 3.2 that the teacher could only resolve the paradox by altering the definition of "surprise", so the students are actually correct that the teacher could not fully fulfill the requirements). As a result, this paradox could be easily resolved. Specifically, the teacher could simply give the test on any random given day. This will, of course, surprise the students as they directly ruled out the possibility of such a test occurring. This argument is also proposed by Quine in his renowned paper “On a So-called Paradox”, in which he argues that a test on any day would satisfy the conditions and would be unforeseen [5]. Here, an argument belonging to View 2 could help explain the negation of View 1. There is a major flaw in the student’s reasoning, which is that the student only considers a scenario in which (s)he believes in the announcement (the student could only rule out Friday to Monday under the assumption that the announcement is true). In other words, the students’ reasoning partly depends on the premise that the announcement is true, and that they believe the announcement is true. Yet, because in reality the student actually comes to the conclusion that (s)he does not believe in the announcement, the line of reasoning is using a false premise, and contradicts itself (it uses the fact that the students believe in the announcement and results in the fact that the students disbelieve in it). The student’s argument is actually Reductio ad Absurdum and is incoherent (as the student would actually perceive the announcement to be false). This is also what allows the teacher to easily fulfill the announcement and gave a surprising test to the students on Wednesday. Thus, no paradox is present, and this negates View 1.

There is another interesting scenario if we assume that there is a test in the next week (which is the actual truth), and if we assume that the students believe this fact even though their reasoning suggests the contrary. For example, we could assume that a school policy is in place which guarantees that there will be a test in the following week. Then I claim that View 1 is also false if we take that the student does not believe in the announcement (in other words, the students think that the announcement is false, but a test will still happen). I have previously explained that the announcement is composed of two parts: that the test will happen, and that the test will be surprising. In this case, the student takes it for granted that there will be a test; thus, in this case the students view the announcement as false based on the second component – they believe that the test must be unsurprising. However, if the student believes that the test must be unsurprising, then it would only hold true if the test is on Friday (note here that in the paradox, the student does not believe that the test will occur. Thus, this section is actually an extension to the paradox by discussing and considering a more complex aspect of the scenario).

To explain this assertion, let us first consider if the test is on Friday. Then the test will be perfectly unsurprising, since the student is certain that the test will occur on that day. Now let us assume that the exam occurs on Thursday. Then on Wednesday night, the student will know that the test will occur on Thursday or Friday. Yet, it is perfectly acceptable that the test occurs on Friday, since the students believe that the test must be unsurprising (it is actually unsurprising, because by Thursday night the student will know that the test will happen, and it could only happen on Friday, since it is the only day left). Therefore, the student cannot deduce with certainty that there will be a test on Thursday, which, according to the definition of “surprising”, makes a test on Thursday surprising. This does not match the expectations of the student (which is that the test must be unsurprising), which lets the student rule out the possibility of a test happening on Thursday. A similar line of reasoning could be applied to the remaining three days, which results in the claim in the previous paragraph. The specialty of the last day (as the student could rule out all of the other four possibilities) makes it the only day where the student would receive an unfair test. The remaining four days, including Wednesday, would surprise the student.

Therefore, in this case, the student will be surprised if the test is on Wednesday. So View 1 is false. Now, I have discussed the cases where 1) the student does not believe that there will be a test, and 2) the student does not believe that the test will be surprising. Combined, these cases negate View 1, which leads us to the conclusion that the teacher is not contradicting himself, and it is not acceptable to simply claim that the paradox is resolved because the teacher is lying.

Next, I will point out other flaws in the paradox and address the definition of “surprising” in order to fully resolve the paradox.

### 3.2 Part Two

Previously, I have mentioned that the definition of surprise is “on any day, the student cannot deduce for certain that the test will happen on the next day based on his memory (whether the test occurred in previous days) and the information in the announcement”. However, now, as I provide you with the above analysis, it seems relevant that I should broaden the definition in order to validate my logic. I propose that we could redefine the term “surprising” to mean: “on any day, the student cannot deduce for certain that the test will happen on the next day, or that the student does not anticipate the test to be on the next day”. In other words, the test is surprising when students think (according to the previous reasoning) that the test will not take place, but it still takes place. This is precisely the definition that the teacher uses in order to fulfil his promise in the announcement. I thus bring out the flaws in this paradox.

This new definition of “surprising” could effectively bring

ISSN 2959-6149

out the teacher's actual inability to fulfil the paradox. According to the announcement, "surprising" only includes my former definition; in other words, a test is surprising if and only if the student cannot be certain that the test will not occur on the next day. However, the teacher's test on Wednesday is claimed to be "surprising", because actually the "surprising" test is due to the students deducing for certain that the test will not happen on Wednesday, but the teacher actually putting the test on that day. The teacher is using the definition of "surprising" that states "a test is surprising if it does not match the student's expectations", which deviates from the original definition of "surprising" that is provided in the paradox itself. Thus, we actually have a reason to say that the teacher cannot actually fulfil the promise in the announcement if we adopt a rigorous definition of terms.

We can delve deeper and analyze the relationship between these two definitions. The definition given by the teacher in the announcement can be denoted as Def A, while the actual definition that the teacher uses can be denoted by Def B. Def A states that "a test is surprising if and only if the students cannot deduce for certain that the test will occur on the next day". Def B states that "a test is surprising if and only if the students do not expect the test to take place". It is evident that Def A primarily relies on logical reasoning, while Def B mainly discusses the subjective viewpoints of the student about the test. This then raises an intriguing question: does the objective or the subjective matter more when it comes to elements of surprise? Here, I argue that the latter is more likely; surprise is an emotion that you feel when something unexpected or unusual happens in our lives, and we care less about whether the event can be scientifically and objectively predicted beforehand.

This definition of "surprising" actually adheres more to our daily life; we call an event surprising if it does not match what we expect would happen, instead of mostly calling events surprising if we could not predict it beforehand. For example, a "surprising test" in real life would be a test that students are not even aware will happen, (or a test that students are certain will not happen), instead of a test that students are not sure whether it will happen or not. The Merriam-Webster Dictionary also defines surprise to be "the feeling caused by something unexpected or unusual" [6]. The teacher is using this definition to surprise the students, which instead should not qualify to be a genuine "surprise" for them.

We can express this paradox through logical symbols and introduce some abbreviations:

1) En: the exam is on the nth day of the week.

2) KnP: on the nth day, the student knows that P.

3) Sn: the exam is on the nth day of the week, and the student does not know in the previous night that the exam is on that day (so Sn is actually equivalent to $En \wedge \neg Kn\text{-}1En$).

(We can also include a fourth, to distinguish between the definitions of "surprising":

4) Sn2: the exam is on the nth day of the week, and the student knows in the previous night that the exam is not on that day)

Thus, the difference between the two definitions become obvious, if we consider the difference between Sn and Sn2:

1. In the first case, the paradox can be expressed as for $n \in \{1,2,3,4,5\}$, $En \rightarrow Kn\text{-}1\neg Sn$.

2. In the second case, the paradox can be expressed as for $n \in \{1,2,3,4,5\}$, $En \rightarrow Kn\text{-}1\neg Sn2$.

There are also worthy extensions to this paradox that we could consider. For example, let us try to extend the paradox and consider another, perhaps more interesting case that the students actually believe for certain that there will be a test (which is also mentioned above but not extensively discussed). For example, this could be achieved by the method mentioned in the first section of implementing a school policy that states there must be a test every week. This could also be achieved by assuming that the teacher is an honest figure that always does what he promises. In this case, the students face a difficult challenge on Sunday evening: on which day will the test actually occur? According to the logic, the student could rule out the possibility that the test would happen on Friday, Thursday, Wednesday, Tuesday, and Monday, respectively. Yet, as the student must choose a day on which the test would occur, it would seem that the student could not fix one of the five days that they believe the test will occur on.

## 4. Literature References

In this section, I will target other viewpoints made by other philosophers regarding this issue, and provide responses that prove either their incomprehensiveness or their logical flaws.

For example, the metaphysician Paul Weiss argues that "the student's argument falsely assumes he knows that the announcement is true. The student can know that the announcement is true after it becomes true – but not before" [2]. This actually belongs to the category of View 2- that the student's reasoning is false. Weiss is arguing that the students cannot predict whether the announcement is true or false before the surprise test actually occurs sometime in the future; thus, the assumption made by the student's argument that the announcement stands true cannot actually be determined. Weiss takes the standpoint that future events could not be actually objectively determined and assessed until the moment when they happen [7].

However, this view is clearly flawed. For one, we could also impose restrictions (for example, the above-mentioned school policy and the extreme honesty and truthfulness of the teacher) to force the student to take a certain stance on the truthfulness of the announcement. Weiss seems to argue that an event in the future can be neither true nor false, because it has not even occurred yet. However, he does not take into account that a prediction with robust supporting evidence or opposing evidence can be assigned a truth value, and it seems very counterintuitive to say that no matter how certain a prediction is, it cannot be assigned a truth value until the event happens. In this specific scenario, the student is able to make the decision based on several factors, such as the policies of the school, the honesty of the teacher, or whether the announcement is logically fluent. With enough justified knowledge, the student is actually able to predict the announcement's truth value.

Furthermore, Weiss's argument seems to claim that the student does not know whether the announcement is true or false. However, I argue here that it is more important to consider the belief of the students, which is a separate concept when compared with the knowledge of the student. While knowledge is based on facts and evidence, belief is often solely based on an individual's thoughts and personal convictions [8]. The belief of the student is more valuable when considering whether a test is surprising, because I have already analyzed that the students feel surprised because they do not expect the results and view it as unusual. In this case, the belief of the student can be easily fixed, because it is based on personal convictions and thus does not heavily rely on supporting evidence.

We can further analyze the role of knowledge in this paradox. Karl Popper has convincingly and successfully argued that knowledge is actually a belief that we perceive to have a high degree of certainty [9]. In other words, knowledge is based on belief, and one cannot know something without believing that it is true. Thus, knowledge is actually a portion of "belief" in that it also adds the requirement that the belief must be able to be proven using sound scientific evidence. Yet, because it is unnecessary to consider whether or not the students' belief can be scientifically proven (due to the above definition of "surprise"), this extra requirement seems unnecessary, and it is enough to only consider the belief of the students. Similarly, in Plato's Meno, Plato analyzes that knowledge is a stronger version of belief that is not swayed easily; for example, if I know that "the road leads to Rome", I am less likely to be disturbed by the fact that the road initially seems to be leading away from Rome, than if I merely believe in the statement [10]. This also leads us to the conclusion that belief should be higher valued in this paradox, since the students would not be able to be surprised by the test if they could know when it will occur. The presence of "knowledge", which implies truth, undermines the element of surprise; instead, surprise could only be caused by a difference in the student's belief and the actual reality.

## 5. Conclusion

In this article, I attempt to give a new explanation of the surprise test paradox by analyzing the flaws when assuming that the teacher is speaking a self-contradictory statement. I also analyzed the definition of "surprising" and gave another explanation of the paradox to revolve around two definitions of the surprising test. I also analyzed and refuted the analysis coming from other philosophers regarding this famous paradox. From this article, I also conclude that the human feeling of surprise is complex and often cannot be summarized by using one single definition. I also analyzed the complex relationship between logical deductions, expectations, and beliefs in this paradox, which plays a profound and significant role in attempting to provide it a satisfying solution.

## References

[1] „Epistemology,“ Encyclopedia Britannica.

[2] Epistemic Paradoxes, Stanford Encyclopedia of Philosophy, 2022.

[3] J. L. Kvanvig, Paradoxes, epistemic. Routledge Encyclopedia of Philosophy, Taylor and Francis, 1998.

[4] J. Earman. A User's Guide to the Surprise Exam Paradoxes. Dept. of History and Philosophy of Science, University of Pittsburgh, 2021.

[5] W. V. Quine. ON a SO-CALLED PARADOX. Mind, 62 (245): 65–67, 1953.

[6] "Surprise." Merriam-Webster Dictionary, 2024.

[7] Weiss, Paul. The Prediction Paradox." Mind, 1952, 1(242): 265–9.

[8] H. Kim. Kant and Fichte on belief and knowledge. Revista De Estud(I)Os Sobre Fichte, 2018, 17.

[9] Karl Popper. Stanford Encyclopedia of Philosophy, 2022.

[10] R. S. Bluck. Plato's meno. Phronesis, 1961, 6 (1–2): 94–101.