The surprise exam paradox: a note on formulating it and a solution to it

Author: Terence Rajivan Edward, University of Manchester.

Abstract. Some formulations of the surprise paradox involve a pair of unnecessary and controversial assumptions. After identifying these assumptions, I propose a solution to the paradox.

In this paper, I identify and call into question a pair of assumptions that appear in some formulations of the surprise exam paradox. Below is an example of such a formulation, from mathematician Timothy Chow:

> A teacher announces in class that an examination will be held on some day during the following week, and moreover that the examination will be a surprise. The students argue that a surprise exam cannot occur. For suppose the exam were on the last day of the week. Then on the previous night the students would be able to predict that the exam would occur on the following day, and the exam would not be a surprise. So it is impossible for an exam to occur on the last day. But then a surprise exam cannot occur on the penultimate day, either, for in that case the students, knowing that the last day is an impossible day for a surprise exam, would be able to predict on the night before the exam that the exam would occur on the following day. Similarly, the students argue that a surprise exam cannot occur on any other day of the week either. Confident in this conclusion, they are of course totally surprised when the exam occurs (on Wednesday, say). The announcement is vindicated after all. Where did the students' reasoning go wrong? (1998: 41)

Towards the end of this passage, Chow makes the following assumptions. First, if the teacher's students are surprised by the exam, then the teacher has succeeded in giving a surprise exam. Second, if the teacher's students are surprised by the exam, having accepted their argument against the possibility of a surprise exam, then there is a problem with this argument. The same assumptions can be found in various other statements of the paradox (Butler and Chapman 1965: 424; Sorensen 1986: 337; Goldstein 1993: 93; Hall 1999: 647-648; Gerbrandy 2007: 21-22). These assumptions may seem beyond doubt, but in fact it is doubtful that they should be made.

In order to examine the students' argument more carefully, it is useful to interpret it so that it relies on a definition of when the exam would be a surprise. But which definition should we interpret it as working with? In everyday life, it makes sense to say that the teacher has surprised their students if the exam happens on a day that the students have not predicted beforehand, which is why the assumptions above seem beyond doubt. But when discussing the surprise exam paradox, we are not interested in a situation where the students have failed to predict the day of the exam but it was still possible to predict it by deduction. For then the teacher has simply got lucky. They are lucky that they did not have a student who made the deduction.

Given the consideration above, it makes sense to interpret the students' argument so that they are relying on the following definition: the teacher has successfully given a surprise exam if it is impossible for there to be a student in their class who, working only from the officially available information, deduces that the exam will happen on that day. As the term 'deduce' is being used here, to deduce from the officially available information means to reason validly from it. Officially available information includes which day of the week it is, the length of the week, what the teacher said in their announcement that there will be a surprise exam and whether the exam has been given yet or not. I do not think there is anything else that needs to be added to this list here.

Now if we interpret the students as relying on this definition, then they can respond to Chow's statement of the paradox as follows: "Although we did not believe that the exam would occur on Wednesday, and hence were surprised, there could have been a student who only had access to the officially available information yet who deduced that the exam would happen on this day. As we said in our argument, on Tuesday night a student could rule out the later days until only this day was left. The teacher therefore surprised his students by luck, the luck of having no such student. So the teacher has not set a genuine surprise exam, because a genuine surprise exam should be impossible to deduce beforehand from the officially available information." In a real life situation, this response is likely to provoke mockery, because the students ended up surprised. But when philosophers or mathematicians discuss the surprise exam paradox, this response matters. Philosophers and mathematicians are usually interested in a surprise exam that can survive changes in the student body, to include a student who thinks differently.

The student response I have presented reveals that the two assumptions that were identified towards the beginning of this paper are open to doubt. Recall those assumptions: if the teacher's students are surprised by the exam, then the teacher has succeeded in giving a surprise exam; and if the teacher's students are surprised by the exam, having accepted their argument against the possibility of a surprise exam, then there is a problem with this argument. I do not think that the surprise exam paradox should be initially formulated so that the students' surprise is used against them, because that involves making these questionable assumptions.

A solution. In the response from the students that I have presented, they appeal to a hypothetical student. The exam happens on Wednesday and on Tuesday night there supposedly could be a student who deduces this. I shall dispute the consistency of this hypothetical student's beliefs. On Tuesday night, the student reasons that the exam cannot happen on the days after

Wednesday, which means only Wednesday is left, so the exam will happen on this day. However, what is their reason for ruling out the later days? For example, if the end of the school week is Friday, what is their reason for ruling out Friday? Their reason is that there could be a student who comes to school on that day having deduced that the exam will occur on that day and, if that is the case, the exam will not occur on that day, because it will not be a surprise. They rule out Thursday for the same reason. But how then can it be consistent for them to believe that the exam will happen on Wednesday?

Their ruling-out principle is this: the exam will not happen on day X if there could be a student who goes to school on that day having deduced that the exam will occur on that day, working only from the officially available information. Now do they not take themselves to be such a student on Wednesday? Do they not take themselves to be a student who goes to school on Wednesday having deduced that the exam will occur on that day? If so, then it is inconsistent for them to not rule out Wednesday as well, given their ruling-out principle. The hypothetical student cannot consistently believe that the exam will happen on Wednesday.[1]

[1] What though about the case of a student who thinks as follows: (a) they believe that the exam will happen on Wednesday, having reasoned in the way sketched; (b) strangely they do not believe that they are a student who has deduced this by relying only on the officially available information? It is only such a student who can consistently believe that the exam will happen on Wednesday. But when evaluating the surprise exam paradox, we do not want to defend the students' argument by appealing only to this peculiar case. We are interested in whether there can be a student who deduces the day of the exam by relying only on official information while believing that this is what they have done. There are no students who can do this consistently.

References

Butler, R.J. and Chapman, J.M. 1965. On Quine's 'So-Called Paradox'. *Mind* 74: 424-425.

Chow, T.Y. 1998. The Surprise Examination or Unexpected Hanging Paradox. *The American Mathematical Monthly* 105: 41-51.

Gerbrandy, J. 2007. The Surprise Examination in Dynamic Epistemic Logic. *Synthese* 155: 21-33.

Goldstein, L. 1993. Inescapable Surprises and Acquirable Intentions. *Analysis* 53: 93-99.

Hall, N. 1999. How to Set a Surprise Exam. *Mind* 108: 647-703.

Sorensen, R. 1986. Blindspotting and Choice Variations of the Prediction Paradox. *American Philosophical Quarterly* 23: 337-352.