Surprise Exam Paradox

Puzzles and Paradoxes

Rabern

University of Edinburgh

Surprise Exam Paradox. A teacher announces that there will be a surprise exam during the next two days. One student, who is a ideally rational, reasons as follows: "Since the exam will be a surprise, the teacher cannot wait until day 2 to give the exam; because if she does, then on the morning of day 2, I will remember that an exam has not yet occurred, so will know that the exam has to be later on day 2—so it will not be a surprise. Moreover, the teacher cannot give the exam on day 1; because if she does, then on the morning of day 1, I will still know that the teacher cannot wait until day 2 (on the basis of the reasoning I just used), and thus, knowing that an exam has not yet occurred, I will know that the exam has to be later on day 1—so it will not be a surprise. Thus, I conclude that the teacher cannot give a surprise exam." The student is then especially surprised when the teacher gives the exam on day 1! The riddle is: Can the teacher fulfill her announcement? Or was the student's reasoning correct? On the one hand, we have the student's elimination argument, which seems sound. On the other hand, common sense says that surprise exams are possible even when we have had advance warning that one will occur. Thus we have a paradox.

What premises and what principles are relied upon in the student's elimination reasoning? Let's formalise the paradox in terms of epistemic modal logic.[1] We just need a propositional modal language.[2] We will have two operators $\Box_1$ and $\Box_2$:

$\Box_1 A$:= "the student knows on the morning of day 1 that A"

$\Box_2 A$:= "the student knows on the morning of day 2 that A"

The exam can only take place on the afternoon of exactly one of two days: day 1 or day 2. So let's introduce propositions to abbreviate those:

d_1 := "there is an exam on the afternoon of day 1"

d_2 := "there is an exam on the afternoon of day 2"

The teachers announcement is not only that there will be an exam but also that that the exam will be a *surprise*. If an exam is a surprise to a student then the student doesn't know in advance of the exam that it will take place. Thus that at least requires that the student doesn't know on the morning of the exam that the exam will be given later that day.

Definition. An exam on day i is a *surprise* to a student iff the student does not know on the morning of i that the exam will be given on the afternoon of day i.

Since we are modeling knowledge we will make the minimal assumption that knowledge is factive. So assume the following schema is valid (for all i):

(T) $\Box_i A \to A$

A further ingredient is that the student is assumed to be perfectly rational. So they know all logic truths and all the logical consequences of what they know. This is already a background assumption in Kripke semantics for epistemic logic that we are working with.[3]

[1] Here we loosely follow the discussion in Holliday (2016) "Simplifying the Surprise Exam", UC Berkeley Working Paper in Philosophy: https://escholarship.org/uc/item/82w2d085

[2] The language is a multi-modal language with knowledge operators indexed to times. To remind ourselves that we are dealing with epistemic modality, we could instead write K_1 and K_2.

[3] Note: we will only assume K and T, so what we conclude will of course also hold for S4.2 or whatever stronger logic is the logic of knowledge.

With these definitions in place we can start to lay down the premises used in the students reasoning. First consider the the teacher's announcement. They say that there will be a surprise exam over the two up coming days, that is:

(1) $(d_1 \wedge \neg\Box_1 d_1) \vee (d_2 \wedge \neg\Box_2 d_2)$

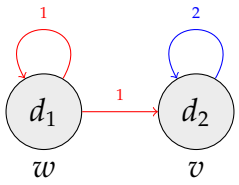
We will also need to make an assumption about the student's memory. If the student wakes up on day 2 and the exam wasn't given on day 1, they will remember this. In other words, if the exam is given on the afternoon of day 2, then the student will know on the morning of day 2 that it was not on day 1 (on the basis of memory).

(2) $d_2 \rightarrow \Box_2 \neg d_1$

The student, then, reasons that given all of this the assumption that there will be an exam leads to a contradiction. That'd be this assumption:

(3) $d_1 \vee d_2$

Are these inconsistent in our modal logic? No. These are consistent given our set up. One can see that (1)-(3) are all true at w in the following Kripke model:



In order to get a contradiction the assumptions need to be stronger. Its not just that (1)-(3) hold but that the student *knows* they hold (and perhaps knows that they will continue to know they hold). It not just that there will be a surprise exam, but that the student knows on the morning of day 1 that there will be a surprise exam.

(1*) $\Box_1((d_1 \wedge \neg\Box_1 d_1) \vee (d_2 \wedge \neg\Box_2 d_2))$

And its not just that the student does in fact remember, but that they know on the morning of day 1 that they will remember.

(2*) $\Box_1(d_2 \rightarrow \Box_2 \neg d_1)$

And finally, assuming there will be an exam, the student knows (on day 1) that it is either on day 1 or day 2, and furthermore they know that they will still know this on day 2. So its enough to show that given the set up the following assumption leads to a contradiction:

(3*) $\Box_1\Box_2(d_1 \vee d_2)$

From these it follows that $\Box_1(d_1 \wedge \neg\Box_1 d_1)$, and by factivity if follows that

$\Box_1 d_1 \wedge \neg\Box_1 d_1$

The culprit here is (1*). The student can't know what the teacher announces (even though it is true).