The surprise exam paradox

Imagine that I begin class with the following announcement:

The Announcement

In addition to the final paper and final exam, we will have one pop quiz (for 99% of your grade) on some class day between now and the end of the semester. (The topic will also be a surprise.) I won't tell you which day I am going to give the exam, but I will tell you this: I will definitely give an exam on one of the remaining class days, and on that day you will have no good reason to believe that it will be on that day, rather than some other. (This is just what it means for it be a surprise exam, of course.)

Having taken a course in Paradoxes, you immediately realize that I have said that I am going to do something impossible, and reply to me as follows:

The Reply

Well, you can't give the exam on the last day of class, since then we would know that, there being no class days remaining, you had to give the exam that day; and in that case it would not be a surprise. So we can safely eliminate the last day of class from the list of possible days on which you can give the exam.

But then you can't give it the second to last day of class, either; for on that day we would know that you couldn't wait till the last day - since then it would not be a surprise - and so we would know that you had to give it that day. But then, of course, it would not be a surprise. So we can also safely eliminate the second-to-last day of the semester as a possible date for the exam.

But then you can’t give it the third to last day ….

(and so on, until all the remaining dates on which I could give such an exam are eliminated.)

Something is wrong with your line of reasoning, since I can clearly give you a surprise exam - but what is it?

It is worth pausing a moment to ask what, exactly, is supposed to be paradoxical about this sort of case. An initial thought is that the case is paradoxical for the following reason: Statements like The Announcement could clearly be true; but the line of reasoning pursued in The Reply seems to show that they could not be true.

The Announcement

In addition to the final paper and final exam, we will have one pop quiz (for 99% of your grade) on some class day between now and the end of the semester. (The topic will also be a surprise.) I won't tell you which day I am going to give the exam, but I will tell you this: I will definitely give an exam on one of the remaining class days, and on that day you will have no good reason to believe that it will be on that day, rather than some other. (This is just what it means for it be a surprise exam, of course.)

The Reply

Well, you can't give the exam on the last day of class, since then we would know that, there being no class days remaining, you had to give the exam that day; and in that case it would not be a surprise. So we can safely eliminate the last day of class from the list of possible days on which you can give the exam.

But then you can't give it the second to last day of class, either; for on that day we would know that you couldn't wait till the last day - since then it would not be a surprise - and so we would know that you had to give it that day. But then, of course, it would not be a surprise. So we can also safely eliminate the second-to-last day of the semester as a possible date for the exam.

But then you can’t give it the third to last day …. (and so on, until all the remaining dates on which I could give such an exam are eliminated.)

Something is wrong with your line of reasoning, since I can clearly give you a surprise exam - but what is it?

It is worth pausing a moment to ask what, exactly, is supposed to be paradoxical about this sort of case. An initial thought is that the case is paradoxical for the following reason: Statements like The Announcement could clearly be true; but the line of reasoning pursued in The Reply seems to show that they could not be true.

To develop this line of thought, it is natural to understand The Reply as giving a reductio of The Announcement. To present that reductio, it will be convenient to adopt some abbreviations.

Let’s suppose that there are 100 days left in the semester. Then let’s say:

E1 = The exam will be on day 1 (and so on, for the other days in the semester).
K(P) = The student knows that P.
SE1 = There will be a surprise exam on day 1 = E1 & ¬K(E1)

We can then begin the reductio of The Announcement as follows:

1. SE1 or SE2 or … SE100. (The Announcement, assumed for reductio)
2. ¬E1 & ¬E2 & … & ¬E99 -> K(E100). (1)
3. ¬E1 & ¬E2 & … & ¬E99 -> ¬ SE100. (2)
4. SE1 or SE2 or … SE99. (1,3)

At this stage - at premise 4 of the reductio - the student has shown that there will no surprise exam on the last day, and that instead the exam must be given on one of days 1-99. But then repetition of the same reasoning can show that the exam will not occur on day 99:

The Announcement

In addition to the final paper and final exam, we will have one pop quiz (for 99% of your grade) on some class day between now and the end of the semester. (The topic will also be a surprise.) I won't tell you which day I am going to give the exam, but I will tell you this: I will definitely give an exam on one of the remaining class days, and on that day you will have no good reason to believe that it will be on that day, rather than some other. (This is just what it means for it be a surprise exam, of course.)

The Reply

Well, you can't give the exam on the last day of class, since then we would know that, there being no class days remaining, you had to give the exam that day; and in that case it would not be a surprise. So we can safely eliminate the last day of class from the list of possible days on which you can give the exam.

But then you can't give it the second to last day of class, either; for on that day we would know that you couldn't wait till the last day - since then it would not be a surprise - and so we would know that you had to give it that day. But then, of course, it would not be a surprise. So we can also safely eliminate the second-to-last day of the semester as a possible date for the exam.

But then you can’t give it the third to last day …. (and so on, until all the remaining dates on which I could give such an exam are eliminated.)

It is worth pausing a moment to ask what, exactly, is supposed to be paradoxical about this sort of case. An initial thought is that the case is paradoxical for the following reason: Statements like The Announcement could clearly be true; but the line of reasoning pursued in The Reply seems to show that they could not be true.

E1 = The exam will be on day 1 (and so on, for the other days in the semester).
K(P) = The student knows that P.
SE1 = There will be a surprise exam on day 1 = E1 & ¬K(E1)

We can then begin the reductio of The Announcement as follows:

1. SE1 or SE2 or … SE100. (The Announcement, assumed for reductio)
2. ¬E1 & ¬E2 & … & ¬E99 -> K(E100). (1)
3. ¬E1 & ¬E2 & … & ¬E99 -> ¬ SE100. (2)
4. SE1 or SE2 or … SE99. (1,3)

At this stage - at premise 4 of the reductio - the student has shown that there will no surprise exam on the last day, and that instead the exam must be given on one of days 1-99. But then repetition of the same reasoning can show that the exam will not occur on day 99:

5. ¬E1 & ¬E2 & … & ¬E98 -> K(E99). (4)
6. ¬E1 & ¬E2 & … & ¬E98 -> ¬ SE99. (5)
7. SE1 or SE2 or … SE98. (4, 6)

And the same reasoning can be used, again and again, to eliminate all of the days of the semester, beginning with day 100, and working back through day 99, day 98, and so on, concluding with the elimination of day 1:

The Announcement

In addition to the final paper and final exam, we will have one pop quiz (for 99% of your grade) on some class day between now and the end of the semester. (The topic will also be a surprise.) I won't tell you which day I am going to give the exam, but I will tell you this: I will definitely give an exam on one of the remaining class days, and on that day you will have no good reason to believe that it will be on that day, rather than some other. (This is just what it means for it be a surprise exam, of course.)

The Reply

Well, you can't give the exam on the last day of class, since then we would know that, there being no class days remaining, you had to give the exam that day; and in that case it would not be a surprise. So we can safely eliminate the last day of class from the list of possible days on which you can give the exam.

But then you can't give it the second to last day of class, either; for on that day we would know that you couldn't wait till the last day - since then it would not be a surprise - and so we would know that you had to give it that day. But then, of course, it would not be a surprise. So we can also safely eliminate the second-to-last day of the semester as a possible date for the exam.

But then you can't give it the third to last day …. (and so on, until all the remaining dates on which I could give such an exam are eliminated.)

1. SE1 or SE2 or … SE100. (The Announcement, assumed for reductio)
2. ¬E1 & ¬E2 & … & ¬E99 -> K(E100). (1)
3. ¬E1 & ¬E2 & … & ¬E99 -> ¬ SE100. (2)
4. SE1 or SE2 or … SE99. (1,3)

At this stage - at premise 4 of the reductio - the student has shown that there will no surprise exam on the last day, and that instead the exam must be given on one of days 1-99. But then repetition of the same reasoning can show that the exam will not occur on day 99:

5. ¬E1 & ¬E2 & … & ¬E98 -> K(E99). (4)
6. ¬E1 & ¬E2 & … & ¬E98 -> ¬ SE99. (5)
7. SE1 or SE2 or … SE98. (4, 6)

And the same reasoning can be used, again and again, to eliminate all of the days of the semester, beginning with day 100, and working back through day 99, day 98, and so on, concluding with the elimination of day 1:

…..

295. E1 or E2. (292, 294)
296. ¬E1 -> K(E2). (295)
297. ¬E1 -> ¬ SE2. (296)
298. E1. (295, 297)
299. KE1. (298)
300. ¬ SE1. (299)

Since we know that there will be exactly one exam, the combination of premises 298 and 300 is enough to show that there will be no surprise exam - which is enough to show that The Announcement is false. So assuming only that The Announcement is true, we can derive its falsity.

The Announcement

In addition to the final paper and final exam, we will have one pop quiz (for 99% of your grade) on some class day between now and the end of the semester. (The topic will also be a surprise.) I won't tell you which day I am going to give the exam, but I will tell you this: I will definitely give an exam on one of the remaining class days, and on that day you will have no good reason to believe that it will be on that day, rather than some other. (This is just what it means for it be a surprise exam, of course.)

The Reply

Well, you can't give the exam on the last day of class, since then we would know that, there being no class days remaining, you had to give the exam that day; and in that case it would not be a surprise. So we can safely eliminate the last day of class from the list of possible days on which you can give the exam.

But then you can't give it the second to last day of class, either; for on that day we would know that you couldn't wait till the last day - since then it would not be a surprise - and so we would know that you had to give it that day. But then, of course, it would not be a surprise. So we can also safely eliminate the second-to-last day of the semester as a possible date for the exam.

But then you can't give it the third to last day …. (and so on, until all the remaining dates on which I could give such an exam are eliminated.)

1. SE1 or SE2 or … SE100. (The Announcement, assumed for reductio)
2. ¬E1 & ¬E2 & … & ¬E99 -> K(E100). (1)
3. ¬E1 & ¬E2 & … & ¬E99 -> ¬ SE100. (2)
4. SE1 or SE2 or … SE99. (1,3)
5. ¬E1 & ¬E2 & … & ¬E98 -> K(E99). (4)
6. ¬E1 & ¬E2 & … & ¬E98 -> ¬ SE99. (5)
7. SE1 or SE2 or … SE98. (4, 6)

…..

295. E1 or E2. (292, 294)
296. ¬E1 -> K(E2). (295)
297. ¬E1 -> ¬ SE2. (296)
298. E1. (295, 297)
299. KE1. (298)
300. ¬ SE1. (299)

Since we know that there will be exactly one exam, the combination of premises 298 and 300 is enough to show that there will be no surprise exam - which is enough to show that The Announcement is false. So assuming only that The Announcement is true, we can derive its falsity.

However, this reductio of The Announcement has a flaw, which can escape notice at first, but, once pointed out, is fairly obvious. Why is, for example, premise 2 supposed to follow from premise 1? The idea is, of course, that if the exam does not occur on days 1-99, the student will know that it will occur on day 100. But **this** certainly does not follow from the fact that there will be a surprise exam on one of days 1-100; if anything it follows from this **plus the fact that the student knows this**. After all, the student will only be in a position to know that there will be an exam on day 100 if the student knows what is said by premise 1.

This indicates that there is no contradiction in supposing that The Announcement could be true; what we seem to have instead is a contradiction in the supposition that **The Announcement could be both true and known to be true**. This should not be surprising; the paradox does not of course show that surprise exams are impossible, but only that surprise exams announced in advance are impossible.

The Announcement

In addition to the final paper and final exam, we will have one pop quiz (for 99% of your grade) on some class day between now and the end of the semester. (The topic will also be a surprise.) I won't tell you which day I am going to give the exam, but I will tell you this: I will definitely give an exam on one of the remaining class days, and on that day you will have no good reason to believe that it will be on that day, rather than some other. (This is just what it means for it be a surprise exam, of course.)

The Reply

Well, you can't give the exam on the last day of class, since then we would know that, there being no class days remaining, you had to give the exam that day; and in that case it would not be a surprise. So we can safely eliminate the last day of class from the list of possible days on which you can give the exam.

But then you can't give it the second to last day of class, either; for on that day we would know that you couldn't wait till the last day - since then it would not be a surprise - and so we would know that you had to give it that day. But then, of course, it would not be a surprise. So we can also safely eliminate the second-to-last day of the semester as a possible date for the exam.

But then you can't give it the third to last day …. (and so on, until all the remaining dates on which I could give such an exam are eliminated.)

This indicates that there is no contradiction in supposing that The Announcement could be true; what we seem to have instead is a contradiction in the supposition that **The Announcement could be both true and known to be true**. This should not be surprising; the paradox does not of course show that surprise exams are impossible, but only that surprise exams announced in advance are impossible.

While this does weaken the intended conclusion of the argument, it does not remove the paradox. After all, we have all been in classes in which the professor announced that there would be a surprise pop quiz several times in the semester - and in which we believed this - and in which we were genuinely surprised by the pop quizzes. So even if we have an argument that The Announcement cannot be known to be true, we seem to have a paradoxical conclusion.

Let's look at how we might construct a reductio of the assumption that the The Announcement can be known to be true:

1. K(SE1 or SE2 or … SE100).	Assumed for reductio
2. K(¬E1 & ¬E2 & … & ¬E99 -> K(E100)).	(1)
3. K(¬E1 & ¬E2 & … & ¬E99 -> ¬ SE100).	(2)
4. K(SE1 or SE2 or … SE99).	(1,3)
5. K(¬E1 & ¬E2 & … & ¬E98 -> K(E99)).	(4)
6. K(¬E1 & ¬E2 & … & ¬E98 -> ¬ SE99).	(5)
7. K(SE1 or SE2 or … SE98).	(4, 6)
……	
295. K(E1 or E2).	(292, 294)
296. K(¬E1 -> K(E2)).	(295)
297. K(¬E1 -> ¬ SE2).	(296)
298. K(E1).	(295, 297)
299. K(¬ SE1).	(298)
300. K(¬SE1 & ¬ SE2 & … ¬SE100).	(299)
C. K(¬SE1 & ¬ SE2 & … ¬SE100) & K(SE1 or SE2 or … SE100).	(1, 300)

The Announcement

In addition to the final paper and final exam, we will have one pop quiz (for 99% of your grade) on some class day between now and the end of the semester. (The topic will also be a surprise.) I won't tell you which day I am going to give the exam, but I will tell you this: I will definitely give an exam on one of the remaining class days, and on that day you will have no good reason to believe that it will be on that day, rather than some other. (This is just what it means for it be a surprise exam, of course.)

The Reply

Well, you can't give the exam on the last day of class, since then we would know that, there being no class days remaining, you had to give the exam that day; and in that case it would not be a surprise. So we can safely eliminate the last day of class from the list of possible days on which you can give the exam.

But then you can't give it the second to last day of class, either; for on that day we would know that you couldn't wait till the last day - since then it would not be a surprise - and so we would know that you had to give it that day. But then, of course, it would not be a surprise. So we can also safely eliminate the second-to-last day of the semester as a possible date for the exam.

But then you can't give it the third to last day …. (and so on, until all the remaining dates on which I could give such an exam are eliminated.)

Let's look at how we might construct a reductio of the assumption that the The Announcement can be known to be true:

1. K(SE1 or SE2 or … SE100).	Assumed for reductio
2. K(¬E1 & ¬E2 & … & ¬E99 -> K(E100)).	(1)
3. K(¬E1 & ¬E2 & … & ¬E99 -> ¬ SE100).	(2)
4. K(SE1 or SE2 or … SE99).	(1,3)
5. K(¬E1 & ¬E2 & … & ¬E98 -> K(E99)).	(4)
6. K(¬E1 & ¬E2 & … & ¬E98 -> ¬ SE99).	(5)
7. K(SE1 or SE2 or … SE98).	(4, 6)
……	
295. K(E1 or E2).	(292, 294)
296. K(¬E1 -> K(E2)).	(295)
297. K(¬E1 -> ¬ SE2).	(296)
298. K(E1).	(295, 297)
299. K(¬ SE1).	(298)
300. K(¬SE1 & ¬ SE2 & … ¬SE100).	(299)
C. K(¬SE1 & ¬ SE2 & … ¬SE100) & K(SE1 or SE2 or … SE100).	(1, 300)

The conclusion of this argument is not an explicit contradiction; what it says is that the student **knows** both of two contradictory propositions. But this does imply a contradiction, given the following very plausible principle:

Knowledge implies truth: If someone knows that P, it is true that P.

Given this, it follows from (C) that:

(¬SE1 & ¬ SE2 & … ¬SE100) & (SE1 or SE2 or … SE100).

Which is a contradiction. Hence it seems that assuming only that The Announcement is known implies a contradiction; from which it follows that The Announcement cannot be known.

The Announcement

In addition to the final paper and final exam, we will have one pop quiz (for 99% of your grade) on some class day between now and the end of the semester. (The topic will also be a surprise.) I won't tell you which day I am going to give the exam, but I will tell you this: I will definitely give an exam on one of the remaining class days, and on that day you will have no good reason to believe that it will be on that day, rather than some other. (This is just what it means for it be a surprise exam, of course.)

The Reply

Well, you can't give the exam on the last day of class, since then we would know that, there being no class days remaining, you had to give the exam that day; and in that case it would not be a surprise. So we can safely eliminate the last day of class from the list of possible days on which you can give the exam.

But then you can't give it the second to last day of class, either; for on that day we would know that you couldn't wait till the last day - since then it would not be a surprise - and so we would know that you had to give it that day. But then, of course, it would not be a surprise. So we can also safely eliminate the second-to-last day of the semester as a possible date for the exam.

But then you can't give it the third to last day …. (and so on, until all the remaining dates on which I could give such an exam are eliminated.)

1. K(SE1 or SE2 or … SE100).	Assumed for reductio
2. K(¬E1 & ¬E2 & … & ¬E99 -> K(E100)).	(1)
3. K(¬E1 & ¬E2 & … & ¬E99 -> ¬ SE100).	(2)
4. K(SE1 or SE2 or … SE99).	(1,3)
5. K(¬E1 & ¬E2 & … & ¬E98 -> K(E99)).	(4)
6. K(¬E1 & ¬E2 & … & ¬E98 -> ¬ SE99).	(5)
7. K(SE1 or SE2 or … SE98).	(4, 6)
……	
295. K(E1 or E2).	(292, 294)
296. K(¬E1 -> K(E2)).	(295)
297. K(¬E1 -> ¬ SE2).	(296)
298. K(E1).	(295, 297)
299. K(¬ SE1).	(298)
300. K(¬SE1 & ¬ SE2 & … ¬SE100).	(299)
C. K(¬SE1 & ¬ SE2 & … ¬SE100) & K(SE1 or SE2 or … SE100).	(1, 300)

However, if we look a bit more closely, we will see that we are in fact assuming something substantial about the student other than that she knows The Announcement to be true.

Look at the move from, for example, premises 1 and 3 to premise 4. What licenses this move?

One might defend it as follows: "Premises 1 and 3 both say something about what the student knows; and so does premise 4. But what the student is said to know in premise 4 is a logical consequence of the things the student is said to know in premises 1 and 3; hence, if premises 1 and 3 really are true, and the student really does know these things, then the student must also know what he is said to know in premise 4 - and so premise 4 must be true."

This sort of argument assumes a very implausible principle about knowledge:

The closure of knowledge under logical consequence

If someone knows that P, and Q is a logical consequence of P, then they also know that Q; i.e., (K(P) & P⊢Q) -> K(Q).

This principle is implausible, because we are not "logically omniscient"; we often do not know logical consequences of things we know, in part because it is often a non-trivial exercise to determine what is a logical consequence of what.

The Announcement

In addition to the final paper and final exam, we will have one pop quiz (for 99% of your grade) on some class day between now and the end of the semester. (The topic will also be a surprise.) I won't tell you which day I am going to give the exam, but I will tell you this: I will definitely give an exam on one of the remaining class days, and on that day you will have no good reason to believe that it will be on that day, rather than some other. (This is just what it means for it be a surprise exam, of course.)

The Reply

Well, you can't give the exam on the last day of class, since then we would know that, there being no class days remaining, you had to give the exam that day; and in that case it would not be a surprise. So we can safely eliminate the last day of class from the list of possible days on which you can give the exam.

But then you can't give it the second to last day of class, either; for on that day we would know that you couldn't wait till the last day - since then it would not be a surprise - and so we would know that you had to give it that day. But then, of course, it would not be a surprise. So we can also safely eliminate the second-to-last day of the semester as a possible date for the exam.

But then you can't give it the third to last day …. (and so on, until all the remaining dates on which I could give such an exam are eliminated.)

1. K(SE1 or SE2 or … SE100).	Assumed for reductio
2. K(¬E1 & ¬E2 & … & ¬E99 -> K(E100)).	(1)
3. K(¬E1 & ¬E2 & … & ¬E99 -> ¬ SE100).	(2)
4. K(SE1 or SE2 or … SE99).	(1,3)
5. K(¬E1 & ¬E2 & … & ¬E98 -> K(E99)).	(4)
6. K(¬E1 & ¬E2 & … & ¬E98 -> ¬ SE99).	(5)
7. K(SE1 or SE2 or … SE98).	(4, 6)
……	
295. K(E1 or E2).	(292, 294)
296. K(¬E1 -> K(E2)).	(295)
297. K(¬E1 -> ¬ SE2).	(296)
298. K(E1).	(295, 297)
299. K(¬ SE1).	(298)
300. K(¬SE1 & ¬ SE2 & … ¬SE100).	(299)
C. K(¬SE1 & ¬ SE2 & … ¬SE100) & K(SE1 or SE2 or … SE100).	(1, 300)

This sort of argument assumes a very implausible principle about knowledge:

The closure of knowledge under logical consequence

If someone knows that P, and Q is a logical consequence of P, then they also know that Q; i.e., (K(P) & P⊢Q) -> K(Q).

This principle is implausible, because we are not "logically omniscient"; we often do not know logical consequences of things we know, in part because it is often a non-trivial exercise to determine what is a logical consequence of what.

However, while it is not plausible to assume that, in general, knowledge is closed under logical consequence, it is not obvious that this is a devastating problem for the attempted reductio of premise 1. After all, the logical deductions which this argument asks the student to carry out are pretty trivial. So maybe we should just stipulate that the student who gives The Reply is sufficiently smart and attentive to see and make all of the relevant logical inferences - that is, all of the inferences required by the reductio argument.

Is this a plausible response to the objection?

Are there any other hidden assumptions being made by this argument?

The Announcement

In addition to the final paper and final exam, we will have one pop quiz (for 99% of your grade) on some class day between now and the end of the semester. (The topic will also be a surprise.) I won't tell you which day I am going to give the exam, but I will tell you this: I will definitely give an exam on one of the remaining class days, and on that day you will have no good reason to believe that it will be on that day, rather than some other. (This is just what it means for it be a surprise exam, of course.)

The Reply

Well, you can't give the exam on the last day of class, since then we would know that, there being no class days remaining, you had to give the exam that day; and in that case it would not be a surprise. So we can safely eliminate the last day of class from the list of possible days on which you can give the exam.

But then you can't give it the second to last day of class, either; for on that day we would know that you couldn't wait till the last day - since then it would not be a surprise - and so we would know that you had to give it that day. But then, of course, it would not be a surprise. So we can also safely eliminate the second-to-last day of the semester as a possible date for the exam.

But then you can't give it the third to last day …. (and so on, until all the remaining dates on which I could give such an exam are eliminated.)

1. K(SE1 or SE2 or … SE100).	Assumed for reductio
2. K(¬E1 & ¬E2 & … & ¬E99 -> K(E100)).	(1)
3. K(¬E1 & ¬E2 & … & ¬E99 -> ¬ SE100).	(2)
4. K(SE1 or SE2 or … SE99).	(1,3)
5. K(¬E1 & ¬E2 & … & ¬E98 -> K(E99)).	(4)
6. K(¬E1 & ¬E2 & … & ¬E98 -> ¬ SE99).	(5)
7. K(SE1 or SE2 or … SE98).	(4, 6)
……	
295. K(E1 or E2).	(292, 294)
296. K(¬E1 -> K(E2)).	(295)
297. K(¬E1 -> ¬ SE2).	(296)
298. K(E1).	(295, 297)
299. K(¬ SE1).	(298)
300. K(¬SE1 & ¬ SE2 & … ¬SE100).	(299)
C. K(¬SE1 & ¬ SE2 & … ¬SE100) & K(SE1 or SE2 or … SE100).	(1, 300)

The closure of knowledge under logical consequence

If someone knows that P, and Q is a logical consequence of P, then they also know that Q; i.e., (K(P) & P⊢Q) -> K(Q).

Are there any other hidden assumptions being made by this argument?

Let's look again at the move from (1) to (2). This move is **not** licensed by our assumption that the student knows the logical consequences of what she knows; after all, (2) but not (1) says that the student **knows something about her own knowledge**.

Assuming that (1) is true and that the student knows the logical consequences of what she knows, what follows is:

2*. K(¬E1 & ¬E2 & … & ¬E99 -> E100).

which differs from (2). What assumption is needed to get from 2* to 2?

It seems that we need to assume two things. First, that

(i) if ¬E1 & ¬E2 & … & ¬E99, then the student knows this;

and,

(ii) that the student knows (i).

Can you see how (2) follows from (1), if we assume (i) and (ii)? Are (i) and (ii) worrying assumptions?

The Announcement

In addition to the final paper and final exam, we will have one pop quiz (for 99% of your grade) on some class day between now and the end of the semester. (The topic will also be a surprise.) I won't tell you which day I am going to give the exam, but I will tell you this: I will definitely give an exam on one of the remaining class days, and on that day you will have no good reason to believe that it will be on that day, rather than some other. (This is just what it means for it be a surprise exam, of course.)

The Reply

Well, you can't give the exam on the last day of class, since then we would know that, there being no class days remaining, you had to give the exam that day; and in that case it would not be a surprise. So we can safely eliminate the last day of class from the list of possible days on which you can give the exam.

But then you can't give it the second to last day of class, either; for on that day we would know that you couldn't wait till the last day - since then it would not be a surprise - and so we would know that you had to give it that day. But then, of course, it would not be a surprise. So we can also safely eliminate the second-to-last day of the semester as a possible date for the exam.

But then you can’t give it the third to last day …. (and so on, until all the remaining dates on which I could give such an exam are eliminated.)

1. K(SE1 or SE2 or … SE100).	Assumed for reductio
2. K(¬E1 & ¬E2 & … & ¬E99 -> K(E100)).	(1)
3. K(¬E1 & ¬E2 & … & ¬E99 -> ¬ SE100).	(2)
4. K(SE1 or SE2 or … SE99).	(1,3)
5. K(¬E1 & ¬E2 & … & ¬E98 -> K(E99)).	(4)
6. K(¬E1 & ¬E2 & … & ¬E98 -> ¬ SE99).	(5)
7. K(SE1 or SE2 or … SE98).	(4, 6)
……	
295. K(E1 or E2).	(292, 294)
296. K(¬E1 -> K(E2)).	(295)
297. K(¬E1 -> ¬ SE2).	(296)
298. K(E1).	(295, 297)
299. K(¬ SE1).	(298)
300. K(¬SE1 & ¬ SE2 & … ¬SE100).	(299)
C. K(¬SE1 & ¬ SE2 & … ¬SE100) & K(SE1 or SE2 or … SE100).	(1, 300)

The closure of knowledge under logical consequence

If someone knows that P, and Q is a logical consequence of P, then they also know that Q; i.e., (K(P) & P⊢Q) -> K(Q).

Summing up: it appears that we have a strong argument for a very surprising conclusion. The conclusion is that, while The Announcement can be true, it cannot be known - at least by someone able and willing to carry out some basic logical inferences, and who knows some elementary facts about her own knowledge of the situation.

This seems pretty close to the advertised paradoxical result: that, if we make some not-unreasonable assumptions about the students in the class, it is impossible to give those students a surprise exam which you announce in advance, so long as the students believe the announcement.

Could this really be true?

One interesting apparent consequence of this argument is that it seems to point toward the existence of truths which can’t, even in principle, be known.

This is a topic to which we will return when we discuss the relationship between truth and provability, and the question of whether there are truths which cannot, even in principle, be proven.

Next time we will discuss a different paradox of knowledge: a paradox which results from the consideration of a combination of knowledge with certain sorts of self-reference.