*Forthcoming in Synthese*

Daniel Immerman

# Question Closure to Solve the Surprise Test

**Abstract:** *This paper offers a new solution to the Surprise Test Paradox. The paradox arises thanks to an ingenious argument that seems to show that surprise tests are impossible. My solution to the paradox states that it relies on a questionable closure principle. This closure principle says that if one knows something and competently deduces something else, one knows the further thing. This principle has been endorsed by John Hawthorne and Timothy Williamson, among others, and I trace its motivation back to work by Alvin Goldman. I provide counterexamples to the principle and explain the flaw in the reasoning of those who defend it.*

**Keywords:** Surprise Test; Closure Principle; Reliabilism

A teacher tells her students that next week's test will be a surprise test, i.e. a test such that the students won't know the morning it occurs that they'll be having a test that day. One student responds that the teacher is mistaken and the test will not take place. He reasons as follows: suppose he hasn't had the test by Friday, the last day of the week. Then on Friday morning he'll know the test will be on Friday so it won't be a surprise. So he won't make it to Friday without having had the test. Suppose he hasn't had the test by Thursday. Then on Thursday morning he'll use the above reasoning to rule out the possibility of making it to Friday without having had the test and thereby come to know the test will be on Thursday, so it won't be a surprise. So he won't make it to Thursday without having had the test. The same reasoning applies to Wednesday, Tuesday, and finally Monday.

But in fact it is the student, not the teacher, who is mistaken. For when she gives the test on Wednesday, the whole class is surprised, including the student. In other words, none of the students knew on Wednesday morning that they'd be having a test on Wednesday.

I have just presented one version of what is known as the surprise test paradox.[1] By 'paradox' I mean a set of propositions, each of which seems plausible, but that are mutually inconsistent. In the paradox just presented, the set is made up of the premises in the student's argument plus the proposition that the student receives a surprise test on Wednesday. The former are each plausible and together entail that he won't have a surprise test; the latter is plausible and entails that he will.

In this paper I will be offering a solution to this paradox. By a *solution* to a version of a paradox, I mean the identification of at least one questionable proposition from amongst the propositions that make up the paradox, plus an explanation of why the questionable proposition(s) seemed plausible.[2] As

[1]For a nearly complete list of papers written on the paradox, see [7, 14].

[2]Some may wish for something stronger than this, namely an identification of a false proposition. But this cannot always be provided. For instance, consider the epistemicist solution to the sorites paradox concerning baldness. The epistemicist will say that one proposition of the form 'if someone with $i$ hairs is bald, then someone with $i+1$ hairs

I am using it, ‘solution’ is a success term; I will use the term ‘approach’ to refer to candidate solutions.

While there is a vast philosophical literature on the surprise test paradox, I will be offering a novel approach to the paradox. Not only that, my approach that has an important upshot regarding debates beyond the surprise test paradox involving *knowledge closure principles*, principles that say if you know some proposition that entails some further proposition and you meet further conditions, then you know the further proposition. In particular, in defending my approach I argue that there is a flaw in a knowledge closure principle defended by John Hawthorne and Timothy Williamson, among others, and also identifiable in the work of Alvin Goldman. I also identify a problem with an argument used to defend these principles.

My paper has two sections. In the first, I introduce the flaw in this knowledge closure principle. In the second, I apply this result to the surprise test paradox.

# 1 Identifying the flaw in knowledge closure

In this section, I will be arguing that there is a flaw in the following knowledge closure principle, which is endorsed by various philosophers, such as John Hawthorne [13, 32] and Timothy Williamson [29, 117]:

*Closure*: If S knows P and S has (competently) deduced Q from

is bald’ is false. But it is part of the epistemicist solution that we do not know which proposition this is.

P then S knows Q.

Why might one endorse this principle? Williamson defends it by saying that it is true 'if deduction is a way of extending one's knowledge' [29, 117] and Hawthorne cites this approvingly [13, 32]. Let us expand on this idea slightly. The basic idea behind this defense it is that the process that ultimately yields the deduced belief Q is a two-step process. The first step in the process is forming the belief that P, the second step is the deduction from P to Q. So long as each step in the process is good enough, the overall process should be as well, for the overall process is a combination of the two steps. In our principle, the first step (the forming of belief in P) results in knowledge and is thus good enough. The second step (the move from P to Q) is competent deduction and thus should also be good enough. So the overall process that leads to belief in Q should be good enough, and thus yield knowledge.

Williamson and Hawthorne are not the first to endorse this sort of argument in defense of this sort of principle. It also appears in Alvin Goldman's article 'What is Justified Belief' in which Goldman first proposed that justification (a condition held to be tightly connected to knowledge [11, 89]) should be understood in terms of reliable belief-forming process. In this article, Goldman makes a distinction between two categories of belief forming processes. The first category is called 'belief-dependent process' and consists of processes that have beliefs among their inputs, such as reasoning and memory. Here is an example: a process that inputs the belief that it is raining and outputs the belief that I should bring an umbrella to work. The second category is called 'belief-independent processes' and includes process that do not have beliefs among their inputs, such as perception. Here is an example: a process that inputs certain visual data and outputs the belief that it is raining.

Because Goldman was a reliabilist, he understood justification (and thus knowledge) in terms of reliability. So for him, a belief-dependent process being good enough was understood in terms of reliability. In particular, it had to be conditionally reliable, where: 'a process is conditionally reliable when a sufficient proportion of its output-beliefs are true given that its input-beliefs are true' [11, 98]. With this out of the way, he proposed the following principle regarding when a belief that resulted from a belief-dependent process was justified:

> If S's belief in p at t results ('immediately') from a belief-dependent process that is (at least) conditionally reliable, and if the beliefs (if any) on which this process operates in producing S's belief in p at t are themselves justified, then S's belief in p at t is justified. [11, 98].

Because the differences between knowledge and justification aren't important for us here, I will alter the wording slightly to talk about knowledge:

> Closure$_{generalized}$: If S knows P and S has formed a belief that Q as a result of a conditionally reliable belief-dependent process starting from S's belief that P, then S knows Q.

As can be checked, Closure is a restriction of Closure$_{generalized}$ to deduction, so long as we understand (competent) deduction in terms of conditional reliability.

In sum, we see Goldman offering what is essentially the same idea as Williamson and Hawthorne. The idea is that if you have a belief, Q, formed by belief-dependent process from P, this belief results from a two-step process. The first step in the process is forming the belief that P, the second step is the move from P to Q. So long as both steps in the process are good enough – here understood as being reliable enough – the overall process will be as well, for the overall process is simply a combination of the two processes. In our principle, the first step (the forming of belief in P) is good enough – it is a case of knowledge, and the second step (the move from P to Q) is good enough – it's conditionally reliable – so the overall process that leads to belief in Q should count as knowledge as well.

While this general principle might seem well-motivated, it is false. To see this, I will consider two arguments against it, one that relies on parallels in the case of testimony and memory, the other of which involves a long sequence of deductions.

Let us start with the parallels of testimony and memory. In order to talk about testimony, we will have to generalize Closure$_{generalized}$ to get a principle that talks about multiple people. Here it is:

> Closure$_{Interpersonal}$: If S knows P and T has formed a belief that Q as a result of a conditionally reliable belief-dependent process

starting from S's belief that P, then T knows Q.

But this principle is false. Getting knowledge from someone via testimony is not simply a matter of them knowing the thing and competently transmitting it. In particular, in order to provide knowledge to someone, the testifier also needs to be disposed to avoid transmitting things that are likely to be false. For example, suppose my friend always tells me (in the same tone of voice) whatever he happens to believe, no matter how likely the proposition in question is to be true. Then even if he happens to tell me something he knows, I will not get knowledge from him. So Closure$_{Interpersonal}$ is false.

What has gone wrong with the argument for Closure$_{Interpersonal}$? The problem is that the belief-forming process that results in my getting testimony from my friend is best described as his coming to believe something and then telling me, not his coming to know something and then telling me. The fact that he happens to know the thing in this case makes no important difference; he would have told me the thing whether it was knowledge or merely something that popped into his head that didn't amount to knowledge. And when we look at the belief-forming process that I'm using to form beliefs, we find that it is unreliable; coming to believe things as a result of him telling me what pops into his head is not a reliable way of forming beliefs. In other words, the overall belief-forming process is not a combination of his coming to know and then conditionally reliably transmitting the knowledge, and this explains how the overall process can be unreliable even if his coming to know something is a reliable process and his transmission is conditionally reliable.

A similar sort of example holds in the case of memory. For example, suppose long ago I came to believe a large number of propositions. Many of these beliefs were false, but some amounted to knowledge. Suppose further that, while I have retained all of these beliefs, I no longer remember how I came to believe them and can notice no epistemically significant difference between them. Then, even if one of these beliefs originally amounted to knowledge, it no longer does. After all, my current belief on the basis of memory is like testimony from a former self; just as I cannot come to know a proposition if I hear it from someone who speaks mostly falsehoods, all in the same tone of voice, likewise I can't come to know a proposition via memory if memory delivers mostly false propositions, with no distinguishable difference between the false propositions and the true ones. So Closure$_{generalized}$ is false.

Here a similar problem has occurred to that in the case of testimony. The problem is that the belief-forming process that results in my remembering the true proposition is best described as my coming to believe something and then remembering it, not my coming to know something and then remembering it. The fact that I happened to know the proposition in this case makes no important difference; I would currently believe the proposition whether it was originally knowledge or merely an idle belief. And when we look at the belief-forming process that I'm using to form beliefs, we find that it is unreliable; believing a proposition as a result of retaining in memory a proposition I formerly believed is not a reliable way of forming beliefs, given that most of the things I once believed are false. In other words, the overall belief-forming process is not a combination of my coming to know and then conditionally reliably retaining the knowledge in memory, and this explains how the overall process can be unreliable even if my coming to know the proposition originally was a reliable process and my remembering is conditionally reliable.

I have just argued that Closure$_{Generalized}$ is false, relying on cases of testimony and memory. I also have a second argument which relies on a case from Joshua Schechter:

> Consider a very long sequence of competently performed simple single-premise deductions, where the conclusion of one deduction is the premise of the next. Suppose that I am justified in believing the initial premise (to a very high degree), but have no other evidence about the intermediate or final conclusions. Suppose that I come to believe the conclusion (to a very high degree) solely on the basis of going through the long deduction. ... it is epistemically irresponsible for me to believe the conclusion. My belief in the conclusion is unjustified. [25].

The explanation I have given for why Closure$_{Generalized}$ is false nicely handles this case. In particular, whether I will decide to deduce, in the final stage of the series of deductions, doesn't depend on whether I knows the conclusion of the previous stage. Even if I had made a mistake at an earlier stage, so long as I am unaware of this, I will still choose to deduce in the final stage. Thus, even though I have been competent the whole way through, and even though I knew the initial premise, this was not enough to yield knowledge of the conclusion.

It should be noted that Schechter's own explanation of his judgment is, at least at first glance, somewhat different from my own – it relies on the claim that I possess a defeater for the claim that the deduction was competently performed [25, 437].[3]

An *(undefeated) defeater* is something such that, if it were not there, one would have knowledge, but because it is there, it keeps one from having knowledge. For example, suppose I read on an assignment sheet that I will be having a test on Friday and come to believe it. Next, a friend who I know to be generally trustworthy tells me that the test has been moved to Thursday. As it happens, the friend is mistaken, and the test is still on Friday. Nonetheless, this testimony forms a defeater for my original belief, robbing me of knowledge.

Schechter says something similar happens in his case: I lack knowledge because there is a defeater for my belief in the conclusion of my deduction. In particular, the defeater arises because it is rational to doubt that my reasoning was fully reliable. This defeater robs me of knowledge of the conclusion of my deduction.

But it is not clear to me how different this explanation really is. The reason that Schechter thinks we have this defeater is because our deductive reasoning isn't fully reliable [25, 440]. So, it is the lack of reliability that leads

[3]Thanks to an anonymous referee for this journal for pressing this worry.

to the defeater that leads to the lack of knowledge. The lack of reliability is what's doing the real work here, in the sense that, whenever one lacks the reliability, one will thereby get the defeater. So I don't think of Schechter as offering a competing explanation – if one wants to say that there is an explanatory intermediary in between the lack of the reliability and the lack of knowledge, viz. a defeater, one can feel free to put it in.

What is it that causes this lack of reliability? Why is it that my belief in the conclusion at the result of this long deduction is not fully reliable. I take it that the natural explanation is the following: while I was unlikely to make a mistake at the last stage of the deduction, I was likely to start that stage with a belief that was false. This isn't because I was likely to start the very beginning of the series of deductions with a false belief, but rather because I was likely to make a mistake somewhere in the middle of the series of deductions and because when I was starting the last deduction, I couldn't tell whether the proposition I was using to deduce the final conclusion were the result of a previous mistake or not. So it seems that in this case too, we have a failure of reliability of one's belief-forming process of the sort I mentioned before. In particular, the cause of the failure of reliability at later stages in the deduction is that one is disposed to deduce conclusions in similar cases even from propositions one doesn't know.

To see this more fully, let us suppose that the deducer Schechter describes has a terrible memory; she starts deducing from a piece of knowledge, and then continues to deduce and deduce, but at any point in the middle of the process, she'll have no idea how long she's been deducing for. Then, it seems to me, she will lack knowledge at every step in the process; it doesn't matter whether she's on the fifth step or the five millionth. Her process at any of those steps can be described as follows: she starts from some proposition she believes and then deduces another from it. She can't tell if the proposition she starts with at this step is known or not; she will deduce regardless. So even if she's only on her fifth deduction, and thus unlikely to have made a mistake by that point, she still lacks knowledge because the process by which she believes the conclusion – deducing from something else she believes – is unreliable. Schechter's own view about defeaters makes this point even clearer – surely even at the fifth step she will possess a defeater, seeing as she's unable to tell how far she is in the process and thus should think it's fairly unlikely that her belief was reliably formed – and thus fail to know.

It might be objected that my explanation of Schechter's original case is not sufficiently general.[4] Take a case in which one attempts to prove a conditional statement via a long conditional proof. In particular, suppose one assumes the antecedent of a conditional, then runs through a long series of deductions until one reaches the consequent, and then discharges the assumption to derive the full conditional. Intuitively, one does not know the conditional at the end of the proof - but it seems that my previous explanation will not handle this case because in offering this sort of proof it is not required that one know the antecedent.

[4]Thanks to an anonymous referee for this journal for pressing this worry.

In response, I first want to point out that this case is not a counterexample to Closure, because, even in the good cases of conditional proof, i.e. those that yield knowledge, one does not know the antecedent of the conditional. Nonetheless, in the good cases of conditional proof, there is some positive epistemic relation one bears to the antecedent and to everything that one deduces from it. Call this positive epistemic relation "the good relation." The corresponding closure principle is the following:

> Closure$_{GoodRelation}$: If S bears the good relation to P and S has formed a belief that Q as a result of a conditionally reliable belief-dependent process from S's bearing the good relation to P, then S bears the good relation to Q.

I can give an analogous version of my explanation of Closure failure to explain why Closure$_{GoodRelation}$ fails. In particular, in a case like the one of the conditional proof, my belief-forming process, near the late stages of the deduction, is unreliable. The failure of reliability seems to be generated not by a failure to bear the good relation to the antecedent, nor a failure to competently deduce any particular stage, but rather by a failure to only make deductions from propositions to which one bears the good relation. In other words, the failure that keeps one from knowing the conclusion of an extremely long conditional proof is the failure to reliably make deductions only from propositions that follow from the antecedent of the conditional. So a generalized version of my explanation can handle this case.

Another objection is that my explanation of Schechter's case might lead to skepticism about mathematics.[5] But I do not think this is so. So long as mathematicians are reliable in their belief-forming practices, I can grant them knowledge. Schechter's case is one involving an unusual sort of unreliability – usually mathematicians do not make extremely long deductions, or, if they do, they check and recheck them.

In sum, I have argued in this section that Closure is false – knowing one proposition and then competently deducing a second is not enough to guarantee knowledge. In particular, one will not gain knowledge if the belief-forming process by which one comes to believe the conclusion is unreliable and if this unreliability stems from a willingness to form deductions even in cases in which one doesn't know the starting proposition in question.

Before I closure this section, I want to consider one objection to what I have said here. It might look as if, in rejecting Closure, I am taking on a large cost because it might appear that many find Closure quite intuitive.[6]

In response: I think that what's intuitive is not that Closure is true, but that some principle in the neighborhood is correct. It is notoriously difficult to correctly state a knowledge closure principle, indeed, so difficult that it is common practice among epistemologists to state a version of a knowledge closure principle, note that it is false, and then say that something in the area is correct.[7] And I have not disputed that something in the neighborhood

[5]Thanks to an anonymous referee for this journal for pressing this worry.
[6]Thanks to an anonymous referee for this journal for pressing this worry.
[7]For examples, see e.g. [1, 37], [4, 12], [9, 64], [14, 207], [15, 40], [23, 245], [27].

of Closure is correct. So I do not have to deny the intuitive thought that motivates Closure.

## 2 Applying the result to the surprise test

In this section, I will apply my result regarding Closure to the surprise test paradox. In particular, recall that the surprise test paradox has the following structure: the student has an ingenious argument that surprise tests are impossible, and yet, when he receives a test on Wednesday it is a surprise; he did not know he would have a test on Wednesday.

Quick preview: on my approach to solving the paradox, the student's argument relies on an instance of Closure that has the flaw that I described in the previous section and this explains why he does not know on Wednesday that he will have a test that day.

In order for this approach to be successful, it must be the case that the student does not know on Wednesday that he will have a test on Wednesday. While most philosophers writing on the surprise test take this for granted, it is surprisingly difficult to defend. I shall now examine and criticize three reasons for thinking that the student doesn't know on Wednesday that he will have the test on Wednesday, and only then offer my preferred explanation of why he doesn't know this. The reason I do this is because my explanation involves some components that I will use to defend my approach to the surprise test. Because I want everyone to accept these components, I want to show that other reasons they might have had for accepting that the student doesn't know he'll have the test on Wednesday are flawed, and thus that if they wish to retain this conclusion, they should accepting my explanation.

One reason for thinking that the student doesn't know on Wednesday that he will receive a test is because it is intuitively obvious that he doesn't know this. Doris Olin expresses this sentiment when she writes:

> ... there is virtually complete agreement that the conclusion of the [student's] argument is false. ... it is pointless to construct further arguments that the surprise examination is possible; this is something we already know. [22, 39].

Olin exaggerates the degree of unity about this argument. While most philosophers seem to think the student's argument is problematic, the same is not true of mathematicians or economists. One group of authors, noting this phenomenon, writes:

> ... of course, the Surprise Examination is treated standardly in the philosophical literature as a *paradox*, thought to hide some fallacious piece of logical legerdemain. That the same form of reasoning is thought of as valid in the theoretical economics literature ... indicates that important work remains to be done in bridging the two bodies of work [8, 163].

Even among philosophers, not all have accepted that the student's argument is unsound. Indeed, the first three philosophers to publish papers on the surprise test paradox all thought the argument was sound [18, 264]. The first of them to publish on the paradox, D. J. O'Connor in 1948, thought it was obvious that the argument was sound, writing, 'It is easy to see that ... [the test] cannot take place' [20, 358].

A second argument offered offer for the claim that the student receives a surprise test on Wednesday is that surprise exams actually occur. After all, real-life teachers often announce surprise exams in advance and then go on to give them.[8]

This argument is problematic because real-life surprise tests do not share some features with the surprise test described in the paradox. Firstly, it is not clear that what is meant by 'surprise test' in real-life cases is the same as what is meant by 'surprise test' in the statement of the paradox. Various commentators have suggested that when real-life teachers announce that there will be a surprise test, by 'surprise test' they mean a test that, if it does not occur on the last possible day, is a surprise.[9] After all, the purpose of a surprise test is not to surprise the students. Rather, the purpose is to make sure the students are prepared. Not knowing which day the test is on, the students will be forced to prepare each night just in case. For example, if a student has not had the test by Tuesday night, and doesn't know if the test will be the next day, she will prepare just in case. But if she knows when the test will be and it's not Wednesday, she might not bother preparing

[8] See, e.g [10, 293].

[9] See e.g. [16, 449-451].

Tuesday night. So the surprise is necessary to get her to prepare early in the week. But suppose she makes it to Thursday night without having had the test. Even if she can deduce that the test will be on Friday, she will prepare anyways. Thus, while having the students not know when the test will be is necessary in order to get them to study early in the week, it is not necessary by the time Thursday night rolls around. So, given the teacher's purposes, what makes the most sense is to announce a test that, if it does not occur on Friday, will be a surprise.

In summary, probably what real-life teachers mean by 'surprise test' – to the extent that they have anything precise in mind – is different from what is meant in statements of the paradox. Furthermore, one cannot run a version of the surprise test argument if 'surprise test' is defined as: a test such that, if it is not given on Friday, the students will not know in advance that it will occur. For if the students make it to Friday morning without having a test, and thus come to know that they will have a test on Friday, it does not follow that they cannot receive a surprise test, on this new definition of 'surprise test.' Thus, the argument the student gives does not apply to real-life surprise tests, and as a result, the fact that they occur does not show that the student's argument is fallacious.

Secondly, in real-life examples of surprise tests, students have not thought of the argument that our student runs through in the statement of the paradox. But it is via the argument that the student takes himself to gain knowledge and thus to avoid being surprised. For example, if the test has not occurred by Wednesday, the student can use the early steps of the argument to rule out the test occurring on Thursday or Friday and infer that it will be on Wednesday. Ordinary students, who have not used this argument, will have no reason to think that the test will be on Wednesday. Of course these real-life students, because they have no reason to think that the test will be on Wednesday, will be surprised when the test occurs on Wednesday. But it does not immediately follow that our student, who is in a different epistemic state from them, would be surprised if it occurred. So even if teachers did mean by 'surprise test' what I mean, the fact that real-life students can know about surprise tests in advance and still be surprised when they occur does not show that our student, who has worked out the argument, is surprised.

A third reason for thinking that the student doesn't know that the test will be on Wednesday is that he has a defeater.[10] Recall that an (undefeated) defeater is something such that, if it were not there, one would have knowledge, but because it is there, it keeps one from having knowledge. Someone offering the third reason says that on Wednesday morning the student has a defeater for his belief that there will be a test on Wednesday morning. The defeater – which the student has good reason to believe – is the following proposition: the student does not know he will have the test on Wednesday. The reason this functions as a defeater is that you can't know something if you have good reason to think you don't know it.

While this reasoning may work for the version of the surprise test paradox

[10]Discussions of this reason include those by [6, 424] [19, 464] and [21, 228-231].

I originally presented, it doesn't work for a variant:

> Conditional Surprise: A teacher announces to her students: 'next week, if I give you a test at all, I will give you a conditional surprise test. By a "conditional surprise test" I mean a test such that, on the day it occurs, you won't know the following material conditional: if you have a test at all that week, you will have a test that day.' The student reasons as before: 'Suppose I make it to Friday without having had a test. Then on Friday morning I will know that if I have a test at all this week, it will occur on Friday. So I won't have a conditional surprise test on Friday. So if I have a test this week, I will not make it to Friday without having had it. Suppose I make it to Thursday without having had a test. Then on Thursday morning, I will use the above reasoning to come to know that if I have a test this week, I won't make it to Friday without having had it. I will thereby come to know that if I have a test at all this week, I will have a test on Thursday. So I won't have a conditional surprise test on Thursday. The same reasoning rules out Wednesday, Tuesday, and finally Monday. Therefore,' the student conclude, 'the test will not occur at all.'
>
> But in fact, it is the student, not the teacher who is mistaken. For when she gives the test on Wednesday the whole class is surprised, including the student. In other words, none of the students knew

> on Wednesday morning that if they had a test at all that week, they would have one on Wednesday.[11]

The reasoning involving a defeater cannot show that in this story, a surprise test occurs. For there are no defeaters of the type in question; this is not a situation in which the student has good reason to believe some proposition and at the same time has good reason to think he doesn't know the proposition. For example, on Wednesday, the student has available an argument for the following statement: if the test occurs at all it will be Wednesday. But he doesn't have good reason to think that he doesn't know this statement. Admittedly, he does have good reason to believe the following: if he's going to have a test that week, then he doesn't know the statement. But this is relatively harmless. It doesn't claim he doesn't know the statement; it only claims that if a certain condition is true, then he doesn't know the statement.

We have just been examining some reasons to reject the claim that the student did in fact know on Wednesday morning that there would be a test on Wednesday. At this point it should be clear that it is harder to argue for the conclusion that the student did not know this than one might have originally anticipated. Nonetheless, there is a pretty good reason for thinking this.

In order to understand this reason, it is important to note the following fact: the student's reasoning, as I presented it, says that for each day of the week, if the student makes it to that day without having had the test,

[11] See [2, 125] and [28, 233] for this formulation of the surprise test paradox.

the student will know the test will occur on that day. So by the time the student makes it to Wednesday without having had the test, this reasoning will already have delivered the wrong result twice, first in that it said that if he made it to Monday without having had the test, he would know that he would have the test Monday, and second in that it said if he made it to Tuesday without having had the test, he would know he would have the test Tuesday.

The unreliability of the student's reasoning process explains why on Wednesday morning he doesn't know that he will have a test on Wednesday. More formally, the following argument can be constructed which shows that the student doesn't know on Wednesday morning that the test would be on Wednesday:

> (1) The reasoning process that yields the conclusion on Wednesday morning that the test will be on Wednesday is the same reasoning process that yields the conclusion on Tuesday morning that the test will be on Tuesday and on Monday morning that the test will be on Monday.
>
> (2) That reasoning process is unreliable.
>
> (3) If a reasoning process is unreliable, then it does not yield knowledge.
>
> Therefore
>
> (4) The student does not know on Wednesday that the test will

be on Wednesday.[12]

Premises (1) - (3) are all quite plausible. (1) is widely supported in the literature. Most authors, when presenting the paradox, do not bother to state the reasoning past Thursday or Wednesday, but rather note that the student can eliminate the other days. Describing the way in which the student can eliminate the other days, authors use such phrases as 'by the same reasoning' [24, 65] [6, 424] [3, 127], 'for exactly the same reason' [12, 647], 'each day of term is eliminated in this way' [26, 382] and 'the students continue in this way' [5, 550].[13]

(2) is obvious – the reasoning process leads to false beliefs on Monday and Tuesday.

(3) is an instance of a more general view about knowledge which runs as follows:

> Reliability: if a belief-forming process is unreliable, it does not yield knowledge.

Reliability is a popular view – almost everyone thinks that, in order for a belief to be knowledge, it has to have been formed by a reliable belief-

[12]It should be noted that several authors have offered arguments similar to the one that I give, although they have not been as fully fleshed out. James Cargile writes that 'the teacher ... should ... [hold] the test not Monday but Tuesday. Then if the students try to claim to have known Monday evening the test would be Tuesday, the teacher can simply point out that similar grounds hadn't been very reliable with respect to Monday [5, 552]. See also [17, 140], [28, 223].

[13]Some authors, instead of using phrases like these, use phrases like 'by similar reasoning.' We can run an altered version of the teacher's argument to account for this by altering (3) to (3*): If a reasoning process and those it is similar to, are unreliable, then they do not produce knowledge.

forming process.[14] (It should be noted that Reliability is merely stating a necessary condition on knowledge; the idea that reliability is necessary and sufficient for knowledge, a view sometimes known as 'reliabilism,' is far more controversial.)[15]

That said, there is one major complaint about Reliability, namely that terms like 'reliable' and 'belief-forming process' are vague, so that while Reliability is true on some senses of 'reliable' and 'belief-forming process' it is not true on others.

My response: while the term 'reliable' may be vague, under almost every way of making it precise, if a belief-forming process led to the wrong beliefs on Monday and Tuesday and the right one on Wednesday, it is unreliable.

Meanwhile, as of now, we need not worry about different ways of individuating 'belief-forming process.' This is because most people writing on

[14]It is worth noting that Reliability is somewhat controversial, in part because Timothy Williamson has used a very similar principle in an argument for a fairly controversial conclusion. In particular, he used it to argue that various mental states, such as feeling cold, are not luminous, that is, are not such that if one instantiates them, one is in a position to know one instantiates them [29]. To those who take Reliability to be implausible I offer the following challenge: do you agree that the student in the Conditional Surprise does not know on Wednesday that if he receives a test at all that week, he'll receive it on Wednesday? If so, give me an argument for why he does not know this that does not make use of Reliability.

[15]It is less clear that Reliability will remain true if it is altered by replacing 'knows' with 'rationally believes' or 'proves.' This is noteworthy because economists and mathematicians often define 'surprise test' in terms of rational belief or proof. One nice thing about my solution to the surprise test paradox is that it can account for why mathematicians and economists tend to think the student's reasoning is sound. In particular, in calling certain premises of the student's argument into question, I make use of the idea that knowledge requires reliability. But if rational belief and proof do not require reliability, then I need not hold that the student's reasoning is unsound if it talks about a surprise test defined in terms of these notions, and thus my approach is successful in solving such versions of the surprise test paradox.

the surprise test paradox seem to agree that the belief-forming process that the student used on Monday and Tuesday is the same as the one he used on Wednesday, as we have just seen while discussing (1).[16]

I have just argued that the student's belief-forming process by which he forms the belief that the test will be on Wednesday is unreliable. But the student has an argument that conclusion that if he makes it to Wednesday without having had the test, he will come to know that the test will be on Wednesday. So it is important to ask: where does the student's argument go awry?

To see the answer, we must first look at the student's argument. The argument runs as follows: if the student makes it to Wednesday without having had the test, he will come to know certain things from which he will deduce, and thereby come to know, that the test will be on Wednesday. But his test is supposed to be a surprise, so he can't know the test will be on Wednesday. Therefore, he can't make it to Wednesday without having had the test.

The part of the reasoning that I think is questionable is the part that goes: 'he will come to know certain things from which he can deduce, and thereby come to know, that the test will be on Wednesday.' More generally, the student's argument seems to be assuming the problematic knowledge closure principle we considered earlier, namely:

[16] I will avoid endorsing a particular view about the best way to individuate belief-forming processes because I wish my argument to be compatible with multiple ways of individuating belief-forming processes.

*Closure*: If S knows P and S has (competently) deduced Q from P then S knows Q.

It is not an idiosyncrasy of my statement of the paradox that there is such a closure principle present in it; every version of the paradox will require the student to reason about what he knows on previous days and thus to include some sort of closure principle.[17]

I think the relevant instance of this Closure Principle has the same flaw I defended in the previous section. Recall that the flaw is the following: one will not gain knowledge if one is unreliable thanks to one's disposition to deduce conclusions in cases in which one doesn't know the starting proposition in question.

As I just argued, the reasoning process that yields the conclusion that the student will have the test on Wednesday is unreliable; it delivers flawed results on Monday and Tuesday.

Furthermore, it is clear that the student is disposed to deduce conclusions using this reasoning process even from propositions that he does not know. As I just noted, the student's reasoning process by which he comes to believe

[17] Of course, I am assuming every version of the surprise test paradox will involve multiple days on which the test can be given. Some contend that there is a one-day version of the surprise test paradox in which the teacher tells the student: 'tomorrow you will receive a surprise test' and the student nonetheless receives a surprise test. My solution cannot handle this formulation, but only multiple day formulations. That said, I would be in good company if I decided to declare the one-day version not a real version of the surprise test paradox. Timothy Chow, who maintains what is arguably the definitive bibliography on the surprise test paradox, notes that in updating his bibliography, 'I decided that [the one day version] was distinct from the surprise examination paradox, and I even went back and deleted some entries that I had included in previous versions of my bibliography that were concerned only with [it]' [7, 14].

on Wednesday morning that the test is on Wednesday is the same reasoning process by which he comes to believe on Monday morning that the test is on Monday. Meanwhile, the reasoning process by which he comes to believe on Monday morning that the test is Monday includes the premise that if he makes it to Wednesday without having had the test, he'll know Wednesday morning that the test is on Wednesday. The latter premise is false as I argued earlier in this section. So what we have here is someone who is a competent deducer, but who deduces conclusions from propositions that are not known. This is exactly the sort of example that makes Closure false.

In short, the instance of Closure invoked in the surprise test paradox has the same flaw as the general flaw I described in the previous section. So we can reject Closure to solve the surprise test.[18]

## References


[1] Philip Atkins and Ian Nance. A problem for the closure argument. *International Journal for the Study of Skepticism*, 4(1):36–49, 2014.

[2] A. J. Ayer. On a supposed antimony. *Mind, New Series*, 82(325):125–6, 1973.

[3] Robert Binkley. The surprise examination in modal logic. *The Journal of Philosophy*, 65(5):127–36, 1968.



[18] Thanks for comments to Katie Finley, Ryan Hammond, Richard Kim, Graham Leach-Krouse, Marc Moffett, Ted Shear, Jeff Speaks, Fritz Warfield, an audience at the 2013 Pacific APA and two anonymous referees.

[4] Joseph Keim Campbell, Michael O'Rourke, and Harry S. Silverstein. Introduction: Framing knowledge and skepticism. In Joseph Keim Campbell, Michael O'Rourke, and Harry S. Silverstein, editors, *Knowledge and skepticism*, pages 1–24. The MIT Press, 2010.

[5] James Cargile. The surprise test paradox. *The Journal of Philosophy*, 64(18):550–563, 1967.

[6] J. M. Chapman and R. J. Butler. On quine's 'so-called paradox'. *Mind*, 74(295):424–425, 1965.

[7] Timothy Y. Chow. The surprise examination or unexpected hanging paradox. www-math.mit.edu/∼tchow/unexpected.pdf, 2011.

[8] Patrick Grim et. al. *The Philosophical Computer: Exploratory Essays in Philosophical Computer Modeling*. The MIT Press, Cambridge, MA, 1998.

[9] Richard Feldman. Skeptical problems, contextualist solutions. *Philosophical Studies*, 103:61–85, 2001.

[10] Kenneth G. Ferguson. Equivocation in the surprise exam paradox. *The Southern Journal of Philosophy*, XXIX(3):291–302, 1991.

[11] Alvin Goldman. What is justified belief? In George Pappas, editor, *Justification and Knowledge*. Dordrecht, 1979.

[12] Ned Hall. How to set a surprise exam. *Mind, New Series*, 108(432):647–703, 1999.

[13] John Hawthorne. *Knowledge and Lotteries*. Oxford University Press, 2004.

[14] Jonathan Kvanvig. Nozickean epistemology and the value of knowledge. *Philosophical Issues*, 14:201–218, 2004.

[15] Adam Leite. How to take skepticism seriously. *Philosophical Studies*, 148:39–60, 2010.

[16] Don S. Levi. Surprise! *Southern Journal of Philosophy*, 38:447–464, 2000.

[17] Ken Levy. The solution to the surprise exam paradox. *The Southern Journal of Philosophy*, XLVII:131–158, 2009.

[18] Avishai Margalit and Maya Bar-Hillel. Expecting the unexpected. *Philosophia*, 13:263–288, 1983.

[19] T. H. O'Beirne. Can the unexpected never happen. *New Scientist*, 10:464–5, 1961.

[20] D. J. O'Connor. Pragmatic paradoxes. *Mind, New Series*, 57(227):358–9, 1948.

[21] Doris Olin. The prediction paradox resolved. *Philosophical Studies*, 44(2):225–233, 1983.

[22] Doris Olin. *Paradox*. Acumen, 2003.

[23] Duncan Pritchard. Recent work on radical skepticism. *American Philosophical Quarterly*, 39(3):215–257, 2002.

[24] W. V. O. Quine. On a so-called paradox. *Mind*, 62:65–67, 1953.

[25] Joshua Schechter. Rational self-doubt and the failure of closure. *Philosophical Studies*, 163:429–452, 2013.

[26] R. Shaw. The paradox of the unexpected examination. *Mind*, 67:382–384, 1958.

[27] Matthias Steup. Epistemology. *Stanford Encyclopedia of Philosophy*, Winter 2013 Edition.

[28] Timothy Williamson. Inexact knowledge. *Mind, New Series*, 101(402):217–242, April 1992.

[29] Timothy Williamson. *Knowledge and its Limits*. Oxford University Press, 2000.