JOURNAL OF PHILOSOPHICAL RESEARCH
VOLUME 32, 2007

THE SURPRISE EXAM PARADOX: DISENTANGLING TWO *REDUCTIOS*

JOHN N. WILLIAMS
SINGAPORE MANAGEMENT UNIVERSITY

ABSTRACT: One tradition of solving the surprise exam paradox, started by Robert Binkley and continued by Doris Olin, Roy Sorensen and Jelle Gerbrandy, construes surprise epistemically and relies upon the oddity of propositions akin to G. E. Moore's paradoxical '*p* and I don't believe that *p*.' Here I argue for an analysis that evolves from Olin's. My analysis is different from hers or indeed any of those in the tradition because it explicitly recognizes that there are two distinct *reductios* at work in the student's paradoxical argument against the teacher. The *weak reductio* is easy to fault. Its invalidity determines the structure of the *strong reductio*, so-called because it is more difficult to refute, but ultimately unsound because of reasons associated with Moore-paradoxicality. Previous commentators have not always appreciated this difference, with the result that the strong *reductio* is not addressed, or the response to the weak *reductio* is superfluous. This is one reason why other analyses in the tradition are vulnerable to objections to which mine is not.

I. INTRODUCTION

A teacher announces:

> There will be an exam some weekday of next week but you will not be justified in believing what day it will be before the close of the previous day.

The smartest student objects:

> But the exam, as announced, cannot be on Friday. For if it were on Friday then by the close of Thursday I would justifiably believe that the exam is the next

> day, in contradiction of the announcement. But having ruled out Friday, the exam, as announced, cannot be on Thursday either. For if it were on Thursday, then by the close of Wednesday and knowing that the exam cannot be on Friday, I would justifiably believe that the exam is the next day, in contradiction of the announcement.

He then iterates this argument in backward calendar order to rule out the remaining weekdays, to conclude that the exam, as announced, cannot be given. Paradox arises because although the student's argument appears sound, it is possible for the announcement to be true.[1] Were the teacher to give the exam on Tuesday morning then the student would be surprised in the sense that he would not have been justified in believing, by the close of Monday, that the exam is the next day. So the paradox must be solved by showing why, despite appearances, the student's argument is unsound.

In §II, I discuss the connection between his argument and Moore's paradoxical proposition, of the form *p and I do not believe that p*. I argue that surprise must be elucidated either as the failure to be justified in belief or as the absence of foreknowledge. In §III, I show that commentators who adopt the former elucidation typically, and understandably, present the student's argument as a *reductio* for the impossibility of the truth of the announcement. I then show that this *reductio* is *weak* in the sense that it is easily shown to be invalid. In §IV, I argue that this shows that the student can only be charitably taken as offering a *reductio* that is *strong*, in the sense that it is harder to fault, for the impossibility of being justified, when the announcement is made, in holding a true belief in the announcement. This is an important point. Previous commentators have not always appreciated the difference between the conclusions of these two *reductios*, with the result that the strong *reductio* is not addressed, or the response to the weak *reductio* is superfluous. In §V, I give two reasons why it is impossible for anyone to be justified in believing a *Moorish* proposition of the form *p and I am not justified in believing that p*. In §VI, I use this result to show that the student is correct to claim that the exam, as announced, cannot be on Friday, *if* he continues to be justified in believing the announcement by the close of Thursday. Nonetheless the student can still be justified in believing the announcement *when it is made*. So the student is wrong to assume that if he is justified in believing the announcement when it is made then he will continue to be justified in believing it by the close of Thursday if no exam has yet been given. Since his strong *reductio* relies upon this false assumption, it is unsound.

In §VII, I turn to the other version of the paradox in which surprise is taken as the absence of foreknowledge that the exam will be given on a particular day. I show that the way in which commentators typically present the student's argument is as a *reductio* for the impossibility of the truth of the announcement. I then show that this *reductio* is *weak* in the sense that it is easily shown to be invalid. In §VIII, I argue that this shows that the student can only be charitably taken as offering a *reductio* that is *strong* in the sense that it is harder to fault, for the impossibility of his knowing,

when the announcement is made, that it is true. Again, not all commentators have appreciated this point. In §IX, I show that the student is correct to claim that the exam, as announced, cannot be on Friday, *if* he continues until the close of Thursday to know that the announcement is true. This is because it is impossible for anyone to know a *knowledge-blindspot*—a proposition of the form *p and I do not know that p*. Nonetheless the student can still know, when the announcement is made, that it is true. So the student is wrong to assume that if he knows when the announcement is made, that it is true, then he will continue to know, by the close of Thursday, that it is true, if no exam has yet been given. Since his strong *reductio* relies upon this false assumption, it is unsound.

My diagnosis of the fault in the weak *reductio* is reminiscent of Quine (1953). My analysis of the justified-belief paradox is indebted to Doris Olin's insightful solution (1983; 1984; 2003), one that turns on the impossibility of anyone being justified in believing a Moorish proposition. Like my analysis, Roy Sorensen's (1984; 1986; 1988) and Jelle Gerbrandy's (1999; 2007) solutions to the knowledge-paradox, turn on the impossibility of anyone knowing a knowledge-blindspot. Nonetheless, my analysis is significantly different from any of these, as I show in §X and §XI. Far from having reinvented the wheel, these analyses are vulnerable to objections to which mine is not.

II. THE SURPRISE EXAM PARADOX AND MOORE'S PARADOX

Binkley observes that formulations of the paradox that make no use of epistemological concepts, such as diagnoses in terms of self-reference, miss the real nub of the difficulty, which is that the student's apparently flawless argument is "somehow rudely brought to nothing by the actual occurrence of the promised, but apparently impossible, examination" (1968, 128). Binkley was the first to notice a connection with Moore's paradox, namely that if we take surprise as the absence of belief,[2] then in a one-day case the announcement becomes

> There is an exam tomorrow but you don't believe that there is an exam tomorrow.

If the student believes this then he believes that

> There is an exam tomorrow but I do not believe that there is an exam tomorrow.

This has the same logical form as G. E. Moore's (1942, 543)

> I went to the pictures last Tuesday but I don't believe that I did

namely the *Moorean*

> *p* and I do not believe that *p*.

What Binkley did not see is that elucidating surprise as the absence of belief fails to generate paradox. In a five-day version the teacher's announcement becomes

> There will be an exam some weekday of next week but you will not believe that it will be that day before the close of the previous day.

No paradox arises because it is too easy for the student to falsify this prediction about what psychological state he will be in. For example, if the student is irrational enough to believe on *any* day that the exam will be the next day, then it follows that the exam, as announced, cannot be given. But that is not surprising, let alone paradoxical. To ensure paradox we must elucidate the student's being surprised by the exam as his failing to be justified before the day of the exam, in believing that it will be the next day (Olin, 1983; 1984; 2003) or alternatively elucidate it as the student's absence of foreknowledge of the exam day (Sorensen, 1984; 1986; 1988). Then in a one-day case, the respective announcements become, from the student's perspective,

> There is an exam tomorrow but I am not justified in believing that there is an exam tomorrow

of the form

> *p* and I am not justified in believing that *p*

and

> There is an exam tomorrow but I do not know that there is an exam tomorrow

of the form

> *p* and I do not know that *p*

both of which resemble Moore's paradoxical proposition. Call the first form of proposition *Moorish* (as opposed to *Moorean*) and the second, following Sorensen (1988), the form of a *knowledge-blindspot*.

Since the relation between being justified in believing that *p* and knowing that *p* is far from clear, I first address what I call the *justified belief paradox* (§§III–VI) and then separately address what I call the *knowledge paradox* (§§VII–IX).

III. THE WEAK JUSTIFED-BELIEF-*REDUCTIO*

Most commentators who take surprise as the absence of justified belief present the student's argument, as I did in §I, as a *reductio ad absurdum* of the truth of the announcement. Olin (1984, 181, my italics) is a good example:

> The student objects that *the announcement cannot be satisfied*, reasoning as follows: 'If the examination were held on Friday, then on Thursday evening, realizing that no examination had yet been given, I would reasonably expect it on Friday; hence a Friday examination would not be a surprise. But, if the exam were given on Thursday, then on Wednesday night I would be aware that an examination had not yet been given and, recognizing that it cannot be given on Friday, would expect it on Thursday; so a Thursday examination would not be a surprise. And so on through the remaining days. Consequently, *the surprise examination cannot be given*.'

This seems like an entirely fair and natural way of representing the student's argument. Yet representing it in this way obscures the essential structure of the paradox (and thereby makes it more difficult to loosen the grip that the paradox has on us). For the student's argument can only be charitably understood as being one of two *reductios*, one of which is *weak* in the sense that it is easily refuted and the other *strong*, in the sense that it is harder to fault.

The teacher's announcement is essentially the conjunction

(1) There will be exactly one examination next week

and

(2) You will not be justified in believing before the day of the exam which day the exam is.

The first step of the student's argument is that if the conjunction of (1) and (2) is true and the exam were on Friday then by the close of Thursday he would be justified in believing that the exam is the next day, in virtue of being justified in believing (1), in contradiction of (2). In other words, he supposes for eventual *reductio* of the truth of the announcement that the announcement is true in order to rule out Friday as an exam day.

The first thing we should notice is that it makes no sense to say either that the student is justified or that he is unjustified in supposing something for *reductio*, as opposed to his merely assuming or presuming something that he shouldn't. Of course the student may or may not be justified in what consequences he deduces from his supposition, but that is a different kettle of fish. The second thing we should notice is that in order to eliminate Friday as an exam day the student must make two other assumptions. The first is that by the close of Thursday he would know that no exam has yet been given. The second is that by the close of Thursday he would be justified in believing (1) in virtue of continuing to be justified in believing the conjunction of (1) and (2), in other words, the announcement. The first of these assumptions may be granted.

The second may not. Suppose that the teacher says nothing at all about the exam. Or suppose that she makes the announcement, but the student is asleep. The teacher later sets an exam on Friday. Then (1) and (2) are both true although the student is not justified in believing (1) by the close of Thursday. In other words, the truth of the announcement entails no justification for any of the student's beliefs about it. So the student's initial elimination of Friday is faulty if he bases it on the supposition that the content of the announcement is nothing more than a truth, because this supposition is consistent with the possibility that it is a truth of which he is ignorant.

Understandably, one might now protest that this observation is trivial. One might also protest that this formulation of the student's argument is uncharitable. There is justice in the latter protest. Surely the student does not aim to show that the content of the announcement, even if never announced, is a contradiction. To meet this complaint we might set up the paradox more realistically by stipulating

that the student knows that the teacher has just uttered the words of the announcement. But this will not save the weak *reductio*. For the possibility still remains that he is not justified in believing the words that he knows the teacher has just uttered. Perhaps he does not understand what she has said. Or perhaps he knows that she is a habitual liar. So it still does not follow that if the exam is set on Friday then he will be justified at the close of Thursday in believing that the exam will be set the next day. What is *not at all trivial* is that this fact determines the structure of a *second reductio* that we *must, if we are charitable*, attribute to the student. As far as I know, nobody has noticed this except Mark Sainsbury (1995, 93–96) who offers no solution.[3]

IV. THE STRUCTURE OF THE STRONG JUSTIFIED-BELIEF-*REDUCTIO*

The student got into trouble because his supposition that the announcement is true does not entail the claim that he is justified in believing, when the announcement is made, that it is true. The only way that he can repair his position is to *suppose*, for the sake of his attempt to eliminate Friday as an exam day, that he is justified in believing, when the announcement is made, that it is true.

Note again that it is a category-mistake to ask whether the student is justified in making this kind of supposition. The supposition is consistent with the role of an announcement, which constitutes information as well as truth. Wright and Sudbury (1977, 42) even hold that any adequate solution to the paradox must respect the fact that the announcement may inform the student.

Now the student can be seen more charitably as arguing, not that the content of the announcement is a contradiction in all circumstances, but rather that the truth of the announcement is contradicted by the supposition that he is justified in believing the announcement when it is made.

In other words, the student must now argue that if he is justified in believing, when the announcement is made, that it is true, then if the exam is held on Friday then by the close of Thursday he would still be justified in believing (1), and so would be justified in believing that the exam will be held on Friday, in contradiction of (2). If his argument is sound on all its iterations then it follows that if he is justified in believing the announcement when it is made then no day satisfies (2). In other words, if he is justified in believing the announcement when it is made then the announcement is false.

But now it does *not* follow that *the announcement cannot be true* (in other words that the exam, as announced, is impossible). What follows instead is that the student *cannot be justified, when the announcement is made, in holding the true belief that the announcement is true*. Call this the *strong reductio*.

We have now exchanged one paradox for another, since we may consistently suppose that the student has good inductive evidence for believing the teacher's words as she utters them and that what she says is true. The evidence might be her past record of making the same announcement on many occasions and then

setting an exam on a day secretly decided at random (say by drawing slips of paper marked with the weekdays from a hat), with the result that her students are not justified in believing before the day of the exam which day the exam is. Or it might be simply the fact that the teacher has, on many occasions, always told the truth.

Note that the conclusions of the two *reductios* are radically different. Not all commentators seem to appreciate this fact, as we will see in §X. Appreciating it is crucial. Otherwise we are in danger of giving a response to the weak *reductio* when in fact it is the strong *reductio* in play, with the result that the strong *reductio* is not addressed, or of giving a response to the strong *reductio* when in fact it is the weak *reductio* in play, with the result that the response is superfluous.

V. NOBODY CAN BE JUSTIFIED IN MOORISH BELIEFS: TWO REASONS

If no exam has been given by the close of the penultimate day and the student is still justified in believing the announcement then he would be justified in the Moorish belief that

> There will be an exam tomorrow but I am not justified in believing that there will be an exam tomorrow.

But Olin observes that it is impossible to believing anything of the form:

> *p* and I am not justified in believing that *p*.

Olin's observation is correct, but is not decisively supported by the argument she gives for it (Olin 2004, 54):

> [I]f a person *T* is justified in believing a statement, then he is not (epistemically) blameworthy in believing it. But if *T* is justified in believing that he is not justified in believing *P*, then he would be at fault in believing *P*.

For what strictly follows from the first claim, plus the antecedent of the second, is that *T* is justified in thinking that he is blameworthy in believing that *p*, in other words that he is not blameworthy in thinking that he is to be blamed in believing that *p*. Whether it follows that he really is to be blamed in believing that *p* is less clear. As Claudio De Almeida (2001) points out, justified beliefs may be false, so why can't *T*'s justified belief that he is not justified in believing that *p* likewise be false?

Nonetheless, I endorse Olin's claim that nobody may be justified in Moorish beliefs. Notice that I am not claiming that it is *irrational* for anyone to believe anything Moorish.[4] Whatever we think about the latter claim, I need not establish it here. I have two reasons for endorsing Olin's claim.

My first reason stems from a principle that I see in Goldman (1986, 62). To illustrate what Goldman may have in mind, consider an example adapted from Bergmann (2005, 426). Suppose that you visit a widget factory. You see the widgets on a conveyor through a viewing window. The widgets appear to be red, so you

believe that they are indeed red. Your perceptual apparatus is reliable and you have no reason to think that the conditions of lighting are abnormal. Your belief that the widgets are red seems to be justified. But now suppose that the foreman, whom you know is a reliable and trustworthy authority, tells you that the widgets are illuminated by red light. You are now justified in thinking that your original belief that the widgets are red was formed in an unreliable way, even if for some reason you do not actually form this higher-order belief. It now seems plausible to say that now you are not justified in believing that the widgets are red. This seems to be the correct verdict even if the foreman is unaware that a rare glitch in the system has switched the light back from red to white. In that variant, you are justified in believing what is false—that your original belief that the widgets are red was formed in an unreliable way. Nonetheless, surely you still have no right to believe something on the strength of credentials that you should doubt. That would be like accepting a check on the strength of a signature you should think is forged!

We could formulate this as

Goldman's Principle:

$\forall s \forall p$ (If s is justified in believing that she is not justified in believing that p then she is not justified in believing that p).[5]

I also need

Justification-Distribution:

$\forall s \forall p \forall q$ (If s is justified in believing that (p and q) then (s is justified in believing that p and s is justified in believing that q)).

To illustrate this principle, suppose that you are justified in believing that it is cold in London, and that it is foggy in London (say, on the basis of a reliable weather report). Then you are justified in believing that it is cold in London. And you are also justified in believing that it is foggy in London. Illustrations can't prove a principle, but the fact that it successfully resists our best attempts to find a counterexample is a good reason to adopt it.[6]

Now suppose for *reductio* that you are justified in believing a proposition of the form *p and I am not justified in believing that p*. By Justification-Distribution, you are justified in believing that p. By the same principle, you are also justified in believing that you are not justified in believing that p. So by Goldman's Principle, you are not justified in believing that p. So you are and you are not justified in believing that p. This is a contradiction. Since both principles are necessary truths,[7] it is impossible for you—or anyone else—to be justified in Moorish beliefs.

My second reason for endorsing Olin's claim is that being justified in a Moorish belief would falsify its content. By Justification-Distribution, if you are justified in believing that (p and I am not justified in believing that p) then you are justified in believing that p, in which case the second conjunct of your original belief is false. Since Justification-Distribution is a necessary truth, it is impossible for you—or anyone else—to have a *true justified* Moorish belief. Your being justified in your

belief is what falsifies its content. This means that you are not justified in that belief. Justification for a belief is supposed to make it a better guide to the truth. But in this case the justification guarantees a trip into falsehood. A belief that is internally defective in this way is not one a seeker after justification should have. So if you do have it, you are not justified in having it. Another way of describing the self-destructive character of your Moorish belief is that justification always aims to improve your chances of reaching the truth, but any conceivable justification for your Moorish belief guarantees falsehood, so any 'justification' for your Moorish belief is necessarily no justification.

At this point someone might object.[8] He might say that you are not unjustified in your Moorish belief unless you *recognize* that it is impossible for your belief to be true and justified. He might support this claim by comparing Moorish beliefs with beliefs in necessary falsehoods: surely Hobbes was not unjustified in believing that there is a method of constructing, with compass and straightedge, a square that is guaranteed to have the same area as that of a given circle, because he was not aware that it is impossible for his belief to be true. Likewise, you are unjustified in your Moorish belief only if you recognize that any justification you might conceivably have for it would destroy its truth. But working this out is no mean feat—you must employ Justification-Distribution in order to construct the just-mentioned proof.

I have three replies. Firstly, the objection boils down to the claim that being unaware of your epistemic failings gets you off the hook when accused of lacking justification for your beliefs. But it is implausible to think that the more self-aware you get, the harder it gets for you to be justified in your beliefs. So I hold instead that what makes a Moorish believer unjustified in her belief is the fact that any justification she might conceivably have for it makes it false, whether or not she recognizes this fact.

Secondly, even if Hobbes was justified in believing that there is a way to square the circle, his position is crucially different from that of a Moorish believer. She should not form a belief unless she is justified in it. But being justified in it is what leads her inescapably to falsehood. She has shot herself in the foot. So she is responsible for being in an epistemically bad position in a way that Hobbes is not.

My last reply is decisive. Assume that I am wrong; you are indeed unjustified in your Moorish belief only if you recognize that any justification you might conceivably have for it would destroy its truth. Now suppose that the student is perfectly rational and self-reflective. The justified-belief paradox remains. And in this case the student cannot be justified in believing that

> There will be an exam tomorrow but I am not justified in believing that there will be an exam tomorrow.

Surely we cannot deny that the student is unjustified in believing that something is true if he *knows* that any justification he might conceivably have for his belief would destroy its truth!

VI. WHY THE STRONG JUSTIFIED-BELIEF-*REDUCTIO* IS UNSOUND

This result means that it is impossible for the student to be justified in believing that

> There will an exam tomorrow but I am not justified in believing that there will be an exam tomorrow

despite the fact that this might be true. But if the exam was given on Friday and the student continues to be justified in believing the announcement by the close of Thursday, then he *would* be justified in believing this. So the student is correct to claim that the exam, as announced, cannot be on Friday, *if* he continues to be justified in believing the announcement by the close of Thursday.

However it would be a mistake to conclude that the student cannot be justified in believing the announcement *when it is made*. We may suppose that the student has good inductive evidence for believing the teacher's words *as she utters them*. This supposition is consistent with the impossibility of being justified in believing the announcement when no exam has been set by the close of Thursday.

What follows now is that although the student may be justified in believing the announcement when it is made, he cannot continue to be justified in believing it by the close of Thursday if no exam has yet been given. This is so despite the fact that his memory and access to information remain unchanged. Since the student's strong *reductio* relies crucially upon the false assumption that his justification *would* so continue, his argument is unsound. This means that the student cannot even rule out Friday on the supposition that he is justified in believing the announcement when it is made. In fact if the exam is given on Friday, thus making (1) true, then it is impossible for the student to be justified in believing by the close of Thursday that the exam is the next day, thus making (2) true as well. So the announcement may be true when announced, and remain true throughout the week. But although the student may be justified in thinking, when it is made, that the announcement is true, he cannot be justified, by the close of Thursday, in thinking that it is true.[9]

VII. THE WEAK KNOWLEDGE-*REDUCTIO*

Much the same analysis applies to that version of the paradox in which surprise is defined as the absence of foreknowledge. In this version, the teacher announces

> There will be an exam some weekday of next week but you will not know what day it will be before the close of the previous day.

The teacher's announcement now becomes the conjunction

(1) There will be exactly one examination next week

and

(3) You will not know before the day of the exam which day the exam is.

Sorensen's representation (1984, 126, my italics) of the student's argument is typical of those who take surprise as the absence of foreknowledge:[10]

> A clever student objects that *the test is impossible*. He first notes that the test cannot be given Friday since the students would then know on Thursday evening that the test must be Friday. The test cannot be given on Thursday because the students would then know on Wednesday evening that the test is either Thursday or Friday, and they have already eliminated Friday. In a like manner, the remaining days of the week are eliminated thereby 'proving' that the *test cannot be given*.

This represents the student as attempting a *reductio* that the announcement cannot be true. The first step of his argument is that if the conjunction of (1) and (3) is true and the exam were on Friday then by the close of Thursday he would know, in virtue of knowing (1), that the exam would be the next day, in contradiction of (3). In making this first step, the student must assume that by the close of Thursday he would know (1) in virtue of continuing to know the conjunction of (1) and (3), in other words, the announcement.

Again this weak *reductio* is easily faulted. For the student's supposition that the content of the announcement is nothing more than a truth, is consistent with the possibility that it is a truth of which he is ignorant (Sainsbury 1995, 93). Even if the paradox is set up with the stipulation (Quine 1953, 64) that the student knows that the teacher has just announced the conjunction of (1) and (3), this is still not enough to support the weak *reductio*, because the student may know that the teacher has just uttered the exact words of the announcement without knowing that her words are true. So it still does not follow that if the exam is set on Friday then he will know at the close of Thursday that the exam will be set the next day.

VIII. THE STRUCTURE OF THE STRONG KNOWLEDGE-*REDUCTIO*

For reasons parallel to those given in §IV, the only way in which the student can repair his position is to *suppose that he knows, when the announcement is made, that it is true*, in order to eliminate Friday. The student must now argue that if he knows that the announcement is true when it is made, then if the exam is held on Friday then by the close of Thursday he would still know that (1) is true, and so would know that the exam will be held on Friday, in contradiction of (3). If his argument is sound on all its iterations then it follows that if he knows the announcement when it is made then no day satisfies (3). In other words, if he knows the announcement when it is made then the announcement is false. But again it does *not* follow that *the announcement cannot be true* (in other words that the exam, as announced, is impossible). What follows instead this time is that *when the announcement is made, the student cannot know that it is true.* This is the strong *reductio*, but in terms of knowledge. Note again that its conclusion is radically different from that of the weak *reductio*.[11] We now have a different paradox,

since it seems patently obvious that the student may know that the announcement is true when it is made.

IX. WHY THE STRONG KNOWLEDGE-*REDUCTIO* IS UNSOUND

If the exam were given on Friday and the student continues to know the announcement by the close of Thursday, then he would know that

> There will be an exam tomorrow but I do not know that there will be an exam tomorrow.

But as Sorensen (1988) shows, it is impossible to know any *knowledge-blindspot* of the form

> p and I do not know that p.

Besides the uncontroversial principle that

> ***Knowledge is factive***:
>
> $\forall s \forall p$ (If s knows that p then p)

it is plausible to hold that

> ***Knowledge distributes over conjunction***:
>
> $\forall s \forall p \forall q$ (If s knows that (p and q) then (s knows that p and s knows that q)).

Now suppose that you know that (p and I do not know that p). By the distribution of knowledge over conjunction, you *do* know that p. But by the same principle, you know that you do not know that p, in which case the factivity of knowledge ensures that you *do not* know that p. This is a flat contradiction. So it is impossible for the student to know that

> There will an exam tomorrow but I do not now know that there will be an exam tomorrow

despite the fact that this might be true. But if the exam is given on Friday and the student continued to know the announcement by the close of Thursday, then he *would* know this. So the student is correct to claim that the exam, as announced, cannot be on Friday, *if* he continues to know the announcement by the close of Thursday.

This result should not lead us to think that the student cannot know the announcement when it is made. For we may suppose that the student has good inductive evidence for believing the teacher's words. The evidence might be that the teacher has a long record of making the same announcement and then setting an exam that is not foreknown by her students. This supposition is consistent with the impossibility of knowing the announcement when no exam has been set by the close of Thursday. If knowledge is justified true belief, the student may know that the teacher's words are true as she utters them.

These days it would be more prudent to claim that knowledge is non-Getteristically justified true belief. But the student's justification for believing the

teacher's words does not have to be defective in the way illustrated by Gettieristic examples. In those examples, although someone has a justified true belief, this belief fails to be knowledge because it is true by lucky coincidence. For instance, your belief that there is a sheep in the field may be justified by the fact that the animal you are seeing looks just like a sheep (but which is really a dog disguised in sheep-skin). However your belief is true by the lucky coincidence that when you form it, there happens to be a sheep hidden from view behind a tree. Things need not be like this with the announcement. We may consistently stipulate that what makes the teacher's announcement true is that she has decided to give a surprise exam and that she is unfailingly resolute in executing such decisions. Now add the fact that she invariably announces such decisions to her students, has a long history of doing so and that the students know that her word is to be trusted. This justifies the student in believing that her latest announcement is true. But it is hardly a lucky coincidence that it is true, because part of what makes the latest announcement true, namely the teacher's resolution to execute her decision to hold a surprise exam, is part of her general disposition to unfailingly execute such decisions. This in turn is part of what justifies the student in believing that her present announcement is true.

The possibility that the student knows the announcement when it is made is also permitted by a truth-tracking account of knowledge such as Nozick's (1981): you know that *p* just in case you have a true belief that *p* such that were it the case that not-*p* then you would not believe that *p* (variation) and were it the case that *p* under slightly changed circumstances, you would still believe that *p* (adherence). Such accounts are still live options in the light of Adams and Clarke's (2005) recent defense of them against putative counterexamples (see also Adams 2005). Imagine that the teacher's announcement is true and that the student believes it when it is made. Imagine further that the student has experienced the teacher making many such announcements in the past and that she is either entirely truthful or entirely untruthful with respect to them. Then if her present announcement is false, so were all her past announcements, perhaps because no such announcements were followed by any exam at all. In that case, the student would know that all her past announcements were false and so would not believe her present announcement either. So if her present announcement were false then the student would not believe that it is true (thus satisfying variance). Moreover if the teacher were to make the announcement under slightly changed circumstances (such as writing it on the board rather than uttering it) then the student would still believe that the announcement is true on the strength of her past record of setting surprise exams (thus satisfying adherence). So, in the possible scenario we are imagining, he now knows that the announcement is true.

The knowability of the announcement is also accommodated by the relevant alternatives account of knowledge: you know that *p* just in case your evidence for *p* excludes any relevant possibility that not-*p*. The student's knowledge that the teacher is generally trustworthy and sincere and has a long record of announcing

surprise exams that indeed turn out to be surprising, is enough for him to exclude the possibility that she is lying or joking, at least by ordinary standards of evidence. By stricter standards, it is not enough for him to exclude the possibility that the teacher is really a clone manipulated by an evil demon with the intention of deceiving him into mistakenly thinking that there will be a surprise exam. But that, presumably, is not a relevant alternative.

So although the student may know the announcement when it is made, he cannot continue to know it by the close of Thursday, if no exam has yet been given, despite the fact that his memory and access to information remain unchanged. Since the student's argument relies crucially upon the false assumption that his knowledge *would* so persist throughout the week, his strong *reductio* is again unsound. This means that the student cannot even rule out Friday on the supposition that he knows the announcement when it is made. In fact if the exam is given on Friday, thus making (1) true, then it is impossible for the student to know by the close of Thursday that the exam is the next day, thus making (3) true as well. So the announcement may be true when announced, and remain true throughout the week. But although the student may know, when it is made, that the announcement is true, he cannot know by the close of Thursday that it is true.

In either version of the paradox, what makes it seem plausible for the student to rule out Friday is his claim that by the close of Thursday, he would continue to know or be justified in believing (1) of the announcement (its first part). Once we notice that this claim is based on the assumption that he would continue to know or be justified in believing the *whole* announcement—and that this is impossible—the plausible claim can be seen to be baseless.

It might be objected that to claim that a person loses his justification or knowledge solely because of a change in his temporal position is just as paradoxical as the puzzle it was supposed to solve.[12]

I reply that what generates the surprise exam paradox is the fact that a one-day version of the announcement is a paradoxical proposition, namely one that is Moorish or a knowledge-blindspot. This fact has a paradoxical consequence. But that is what we should expect.[13] Moreover, in solving any deep paradox we should be prepared to give up some of our intuitions. It is better to give up the intuition that a person never loses his justification or knowledge solely because of a change in his temporal position than the intuition that the announcement is knowable.

I have now established all I need. In either version of the paradox, the weak and strong *reductio* are both unsound. The two versions of the paradox are therefore solved.[14]

X. RELATED ANALYSES: QUINE

I conclude by distinguishing my analysis from others that might seem similar but that in fact are open to objections to which mine is not.

Quine addresses the knowledge-version of the paradox (1953) but frames it in terms of a judge that decrees to a prisoner called *K* that

(4) You will be hanged on one of the following seven days at noon

and

(5) You will not know the day of your execution until the day itself.

According to Quine, *K* foresees two possibilities by the close of the penultimate day: (a) he will already have been hanged, or (b) he will be hanged the next day and he will know this. *K* sees that (b) falsifies (5) of the decree and so adopts (a), leading him to eliminate the last day. But he overlooks two other possibilities: (c) he will not be hanged the next day, in falsification of (5) of the decree, and (d) he will be hanged the next day and he does not know this (1953, 65–66).

Quine's point is that *K* overlooks the fact that (d) is consistent with the decree as well as (a), so *K* is not entitled to eliminate the last day on the supposition that the decree will be fulfilled. In particular, if *K* supposes that the decree will be fulfilled, then the most he may conclude is that either (a) or (d) is true, because that supposition allows him to rule out only (b) and (c).

Quine also remarks (1953, 66) that

> To suppose that the assumption of the fulfillment of the decree eliminates *(d)* is to confuse two things; (i) a hypothesis, by *K* at *t*, that the decree will be fulfilled, and (ii) a hypothesis, by *K* at *t*, that *K* will know at $t + n - 1$ that the decree will be fulfilled.

Here *t* is the time that the decree is issued and *n* is the last of the seven days. So in other words, when the decree is issued, *K* momentarily slips from the supposition that the decree is true to the supposition that he will know that it is true by the close of the penultimate noon. This is why *K* mistakenly eliminates (d).[15]

In the exam version of the paradox, this translates into a distinction between the student's supposition, when the announcement is made, that it is true, as against his supposition, when the announcement is made, that he *will know that it is true by the close of the penultimate day*. Quine's distinction approaches (but does not quite coincide with) my distinction between the student's supposition, when the announcement is made, that it is true, as against his supposition, when the announcement is made, that he *then knows that it is true*.

What Quine does not see is that the second supposition forms the basis for *K*'s strong *reductio*.[16] To evade Quine's objection, *K* would have to suppose that he knows, when it is issued, that the decree will be fulfilled. Then he may argue that if he is hanged on the last day then by the close of the penultimate day he would know that he will be hanged the next day, in contradiction of (5). If his argument is sound on all its iterations then if he knows, when it is issued, that the decree will be fulfilled, then no day satisfies (5). In other words, if his argument is sound on all its iterations then if he knows, when it is issued, that the decree will be fulfilled, then the decree will not be fulfilled. Now it follows, not that that decree cannot be fulfilled, but rather that *K* cannot know, when it is issued, that the decree will be fulfilled.

To fail to address this strong *reductio* invites skepticism about the future or testimony. By contrast, I have offered reasons why *K* (or equivalently, the student

in the exam case) could know this at the start of the week. By my lights, the fault in *K*'s strong *reductio* lies in his false assumption that if he knows, when it is issued, that the decree will be fulfilled, then he will continue to know this at the close of the penultimate day.

XI. RELATED ANALYSES: OLIN

Olin and Sorensen both offer analyses that resemble mine. One big difference is that neither distinguishes the weak from the strong *reductio*. Olin's response to the justified-belief paradox turns on observing that by the close of Thursday the student cannot be justified in believing that

> There is an exam tomorrow but I am not justified in believing that there is an exam tomorrow.

Olin observes that it is impossible for the student to be justified in believing this (2005, 53–54). As we saw in §V, this is correct, although not for the reason that Olin gives. She concludes that the student "cannot justifiably believe the announcement on Thursday night and the argument is blocked at the very first step" (2003, 55; 1984, 182). But what is the conclusion of this argument supposed to be? Olin recognizes that

> [t]he announcing of the surprise examination is crucial to the problem because it must be plausible to suppose, at certain points in the argument, that the student would have good reason to believe that a surprise examination will be given. (Olin 1983, 227; 2003, 39)

Nonetheless she still has the weak *reductio* in mind in concluding that "even assuming the details concerning [the students'] memory and reasoning ability, it is possible for the teacher's announcement to be true" (Olin 1983, 228) and that "the surprise exam is thus possible" (2003, 56). So strictly speaking, as a response to the weak *reductio*, Olin's solution is redundant. Olin gives no reason why the student may be justified in believing the announcement when it is made, with the result that the strong *reductio* remains unaddressed. I do not wish to make too much of this, since my analysis may be seen as evolving from hers, but taking care to identify the two *reductios*. Here is Olin's main claim:

> Even though on Thursday night the student has evidence that strongly confirms [the announcement], sees this, and has no other relevant evidence, he is not justified in believing [the announcement]. For he cannot be warranted in believing [the announcement] without also being justified in believing a statement of the form "P and I am not now justified in believing P." (2003, 54)

I do not dispute this. Indeed my analysis builds upon it. Nonetheless, Olin does not distinguish the student's being justified in believing the announcement when it is made, but losing his justification by the close of Thursday, from his never having had any justification for believing it in the first place.

XII. RELATED ANALYSES: SORENSEN

Sorensen's response to the knowledge-paradox employs the notion of a *conditional knowledge-blindspot*. A proposition of the form

p and I do not know that *p*

is a knowledge-blindspot because, as we saw in §IX, it is an unknowable possible truth. A *conditional knowledge-blindspot* is a conditional the consequent of which is a knowledge-blindspot, such as

If I have been drugged then I have been drugged but I don't know it.

I may know this conditional and I may know its antecedent. But I cannot know the conditional *if I also* know its antecedent. For otherwise, given that

Knowledge is closed over known conditionals

$\forall s \forall p \forall q$ (If *s* knows that (if *p* then *q*) and *s* knows that *p*, then *s* knows that *q*)

then I would have to know that (I have been drugged but I do not know that I have been drugged), which, being a knowledge-blindspot, is impossible. Sorensen now construes the announcement as a conditional knowledge-blindspot. Assuming for simplicity of exposition a two-day case, the announcement becomes

(6) There will be exactly one examination on Thursday or Friday.

and

(3) You will not know before the day of the exam which day the exam is.

This is equivalent to

(7) Either there will be an exam on Thursday but you will not know this on Wednesday or there will be an exam on Friday and you will not know this on Thursday.

From the student's point of view, this in turn is equivalent to

(8) If it is not the case that (there will be an exam on Thursday and I will not know this on Wednesday) then there will be an exam on Friday and I will not know this on Thursday.

Sorensen points out that "nothing prevents" the student from knowing (8), in other words the announcement, when it is made. But at the close of Thursday, if no exam has taken place, he learns that its antecedent is true. So he cannot continue to know the announcement (1984, 132–133; 1986, 133; 1988, 329–330). Otherwise he would have to know that

There is an exam tomorrow but I don't know that there is an exam tomorrow

which, being a knowledge-blindspot, is impossible. This refutes the strong *reductio*.

But before this refutation Sorensen (1986) presents the "clever student" as objecting to the teacher's announcement by giving an argument that concludes that "the surprise test is *impossible*" (1986, 504, my italics). This argument can only be the weak *reductio*. Then Sorensen characterizes this argument as follows:

> given that they *retain their knowledge* of the teacher's announcement, accumulate knowledge of past tenseless days, and make rather straightforward inferences, it is impossible to surprise them with a Friday test. (1986, 505, my italics)

This shows that Sorensen has the strong *reductio* in mind. Having employed his solution in terms of conditional blindspots, he rebuts the weak *reductio* by observing that, "the teacher can fulfill his announcement by giving a Friday test" (1986, 506; 1988, 329). As a response to the weak *reductio*, Sorensen's solution, although ingenious, is redundant. It does not address the stronger argument, which is a *reductio* of the knowability of the announcement when it is made.

Sorensen has shown that the following three statements are incompatible: (i) the student knows the announcement when it is made, (ii) if the student continues to know the announcement when it is made, then he continues to know it on Thursday, and (iii) the exam is on Friday. Sorensen wishes to square (i) with (iii), by rejecting (ii). But given (iii) it is unclear why we should reject (ii) rather than (i). In claiming that the announcement is a conditional blindspot for the student on Thursday, Sorensen tacitly assumes that the announcement itself is not a blindspot for the student, in other words that the student can know it when it is made. But this begs the question against the student's strong *reductio*.

By contrast, my analysis has a number of advantages. Instead of merely noting that nothing prevents the student from knowing the announcement when it is made, I have offered reasons why he might know this, consistently with the impossibility of continuing to do so by the close of Thursday. Moreover, my analysis does not invoke conditional blindspots and so is simpler. This means that I need not assume the controversial principle that knowledge is closed over known conditionals.

Robert Nozick (1981) for one would deny this principle for the reason that it leads to skepticism. A skeptic may argue that I know that if I am sitting at home then I am not a brain floating in a vat. But I do not know that I am not a brain floating in a vat. Given the disputed principle, I do not know that I am sitting at home. Nozick's response is to reject the principle, using his truth-tracking account of knowledge, in particular its variation condition. Since it is a necessary truth that if I am sitting at home then I am not a brain floating in a vat, variation does not apply. In fact I believe this truth and would still do so in the nearest worlds (thus satisfying adherence), so I know it. I am sitting at home (or at least we may suppose) and I believe that I am. In the nearest possible worlds in which I am not sitting at home I do not believe that I am sitting at home (thus satisfying variation). And in the nearest possible worlds in which I am still sitting at home, I still believe that I am (thus satisfying adherence). So I know that I am sitting at home. But in the nearest possible worlds in which I am a brain floating in a vat, I still believe that I am not a brain floating in a vat, so variation fails. Thus I do not know that I am not a brain floating in a vat.

If the rational response to skepticism is to avoid it (as most believe) and if Nozick is correct that avoiding it requires rejecting Sorensen's principle, it follows that the principle fails even as a principle of ideal rationality.

I conclude that my analysis of the surprise exam paradox is an improvement over those of Olin or Sorensen.

Finally, since I have received the criticism that my analysis is essentially the same as Jelle Gerbrandy's,[17] I ought to explain why it isn't. Since setting the record straight will prove a bit taxing, I do so in Appendix II below.[18]

APPENDIX I. DEFANGING SORENSEN'S NON-TEMPORAL VARIANT

In this variant the teacher takes five students, A to E, and puts them in an alphabetical queue, starting with A, so that B can see the back of A, and C can see the backs of A and B and so on. The teacher then announces

> Exactly one student has a star on his back but that student will not be justified in believing this before breaking formation.

This announcement has the form

> Exactly one student has a star on his back
>
> and
>
> The student who has the star on his back is not justified in believing he has the star on his back before breaking formation.

An objector argues

> But E cannot have the star. Otherwise he would justifiably believe that none of the other students have it and so would be justified in believing that he has it. Since D will realize this, he cannot have the star either. Otherwise he would justifiably believe that none of the other three students in front of him have it either and so would be justified in believing that he has it.

The objector then iterates this argument to rule out the remaining students to conclude that the announcement cannot be true. But on breaking formation, one of the students, say B, may discover that he has the star.

My analysis has the resources to defang this variation. Once again this *reductio* against the truth of the announcement is weak, since the mere truth of the announcement is consistent with E's failure to justifiably believe it. The objector must instead suppose that each of the five students is justified in believing the announcement. In other words he must now argue that if *each* student were *justified in believing* that the announcement is true then if E has the star then he would be justified in believing, in virtue of being justified in believing the first conjunct of the announcement, that he has the star, in contradiction of its second conjunct. If his argument is sound on all its iterations then it follows that if each student is justified in believing the announcement then no student satisfies the second conjunct of the announcement. In other words, if each student is justified in believing the announcement then the announcement is false. But now it does *not* follow that *the announcement cannot be true*. What follows instead is that *if each student is justified in believing the*

announcement then the announcement is false. This strong *reductio* must assume that E is justified in believing the announcement, if he has the star. But in that case E would be justified in the Moorish belief that

I have the star but I am not justified in believing that I have the star.

We have already seen that this is impossible. So the strong *reductio* is unsound.

APPENDIX II. RELATED ANALYSES: GERBRANDY

Jelle Gerbrandy (1999, 157–163; see also his 2007 paper which presents essentially the same solution) deals with a version of the paradox in which a player (call him *b*) learns from a quiz master that a series of numbered boxes (say five) will be opened in sequence from box 1 to box 5. The quiz master tells him that

(9) Exactly one of the five boxes contains the money

and that he will win the game (and keep the money) if he knows which box contains the money before opening it. The quiz master then also informs *b* that

(10) You will not know which box contains the money before opening it.

In other words, he informs *b* that he will not win the game. Gerbrandy imagines *b* reasoning as follows:

Main Argument

> Suppose that the money is in the last box. In that case, I would know that the money is in that box at the moment when all the other boxes were opened, and I would win the game. But if this is true, the money cannot be in box 4 either, because I know (now) that the last box is empty, and so, if boxes 1 to 3 were opened, the money had to be in the fourth box, and I would win the game as well. I can repeat this proof for all boxes. But then I have to conclude that all boxes are empty. Therefore the quiz master must be lying to me (1999, 157–158).

b's argument is a *reductio*. From the truth of the announcement, together with some obvious facts, he derives the falsehood of the announcement (1999, 158). Paradox arises because, as Gerbrandy points out, all of what the quiz master has announced may be true (1999, 160). Suppose that the quiz master has only yet announced (9) and the money is in box 4. One may imagine the quiz master muttering (10) to himself out of earshot of *b*. Then *b* has no reason to think that the money is in box 4 before he opens it, so (9) and (10) are both true.

Gerbrandy assumes, as part of the set-up of the paradox, that *b* unproblematically knows that (9) is true. Then *b* argues as follows, as preparation for the main argument above:

Preliminary Argument

1.	The quiz master only announces what he knows	(assume)
2.	The quiz master has just announced (10)	(fact)
3.	The quiz master knows that (10) is true	(1 and 2)

4.	I learn whatever the quiz master announces	(assume)
5.	I have just learned (10)	(2 and 4)
6.	I know whatever I have learned	(assume)
7.	I now know that (10) is true	(5 and 6)

Given this conclusion in step 7, *b* may legitimately argue the

First Step of the Main Argument

Suppose that the money is in box 5. I know that (9) is true (as a given in the set-up of the paradox) and I know that (10) is true (from step 7 of the preliminary argument). So after opening the first four boxes, I would know that

The money is in box 5 but I do not know that the money is in box 5

But this is impossible. So the money cannot be in box 5.

b then legitimately argues the

Second Step of the Main Argument

Suppose that the money is in box 4. I know that (9) is true (as a given in the set-up of the paradox) and I know that (10) is true (from step 7 of the preliminary argument). I also now know that the money is not in box 5 (from the conclusion of the first step of the main argument). So after opening the first three boxes, I would know that

The money is in box 4 and I do not know that the money is in box 4

But this is impossible. So the money cannot be in box 4.

And so on. Gerbrandy notes that both steps are valid. Moreover, according to Gerbrandy, *b* can still properly rule out box 5 as containing the money (but not any other box). This is not because of Step 1 of the main argument (which although valid is unsound, as we will see) but because of line 3 in the preliminary argument. Line 3 tells us that the quiz master knows that (10) is true. He also knows that (9) is true (since he announced (9) and he announces only what he knows (line 1)). In other words, the quiz master knows that the money is in one of the five boxes and that *b* will not know which box contains the money before opening it. It follows that the money cannot be in box 5. Otherwise *b* would know, in virtue of knowing (9), that the money is in box 5 after opening the other four boxes. In that case the quiz master would know what is false, which is impossible. Since *b* can work through this piece of reasoning for himself, he may rule out box 5. In other words, the effect of the announcement is that *b* learns something new, namely that the last box is empty (Gerbrandy 1999, 161).

Gerbrandy thinks that the flaw in *b*'s reasoning lies in his preliminary argument, in particular, line 6. *b* is mistaken in assuming that he knows whatever he learns. For example if you tell me

"Your shoelaces are undone but you don't know that"

I may learn what you have told me. But I cannot come to know what you have told me, because knowledge distributes over conjunction. So in coming to know what you have told me, I would come to know that my shoelaces are undone, which falsifies what you have told me, by falsifying its second conjunct.

According to Gerbrandy, learning (10) also falsifies it. *b* already knew (9), in other words, that the money is in exactly one of the five boxes. At line 3 in his preliminary argument, he came to know that the money is not in box 5. Suppose that in learning (10), *b* comes to know that it is true, in other words, comes to know that he will not know which box contains the money before opening it. Then after opening the first three boxes, *b* will come to know that the money is in box 4, which, as we supposed at the very beginning, it indeed is. But then (10) is false. So in the case we have supposed, in which the money is in box 4, *b* cannot come to know that (10) is true, because this falsifies (10) and knowledge is factive. Following Veltman (1996, 223), this is what Gerbrandy calls an "unsuccessful update":

> Updates with sentences that express lack of information of the agent who learns the sentence are not always successful: an agent can learn that such a sentence is true without coming to [know] that the sentence is true. (1999, 161)[19]

So both of *b*'s steps are unsound because in both, *b* mistakenly assumes that he still knows that (10) is true. Nonetheless *b* may rule out box 5 in his preliminary argument at line 3, for at this point he does not need the false assumption that he knows whatever he learns, but only needs it later, at line 6.

In summary, supposing that the money is in box 4, although the quiz master spoke the truth when he said that *b* could not win in the initial situation, the effect of the announcement is that the situation changes to the effect that *b* will win the game after all (Gerbrandy 1999, 162). In other words, the paradox is solved because there is no real conflict between *b*'s conclusion and the fact that the quiz master's announcement may be true. For once *b* learns (10) it becomes false.[20]

My analysis is radically different in three ways. By my lights, the student cannot even rule out Friday on the supposition that he knows or is justified in believing the announcement when it is made. By contrast, Gerbrandy endorses *b* in ruling out the last box, on the assumption that he takes the quiz master to know the announcement. Secondly, Gerbrandy holds that the announcement is falsified by *b*'s learning it. On my account, the announcement remains true throughout the week but becomes impossible for the student to know or justifiably believe at the close of the penultimate day. Thirdly, I do not set up the paradox by assuming that the student knows the first part of the announcement (that there will be an exam exactly one day next week). By contrast, Gerbrandy sets up the paradox by assuming that *b* knows that the money is in one of the five boxes. And unlike Gerbrandy I do not assume that the teacher (quiz master) only announces what he knows.

Gerbrandy's solution is open to four objections that my analysis is not. Firstly, *b*'s argument, which opens this section and which concludes "the quiz master must be lying to me" is the weak *reductio*. To show that it is unsound, all that is needed is to note that the supposition that (10) is true is consistent with *b*'s not hearing it. So strictly speaking, Gerbrandy's response to it is redundant. Secondly, in order for *b* to rule out the last box in his preliminary argument, Gerbrandy must also assume, not only that the quiz master only announces what he knows, but that *b* knows this as well. So the paradox is set up with the assumption that *b* knows, at the time of the announcement, that the quiz master knows that his announcement is true. But this does not seem to be an essential part of the original paradox because *b* would be able to produce the strong *reductio* without this extra assumption, merely by supposing, for the sake of argument, that he knows that the announcement is true. So Gerbrandy's response does not seem to get to the heart of the original paradox we have been considering. Thirdly, it is one of the platitudes about truth that if a statement is true then it is true forever (Wright, 1995). In particular, Margalit and Bar-Hillel (1983) claim that there can be no alteration in the truth-value of the announcement (but for dissent, see Jongeling and Koetsier, 1993). Finally, to say that an agent can learn that a statement (the announcement) is true without coming to know that it is true is hard to swallow. It is part of the very nature of learning that if you learn that *p* then you come to know that *p*. Otherwise there would be no contradiction in saying, "I've learned a lot today but I'm just as ignorant as I was yesterday." I admit that the case in which you tell me,

"Your shoelaces are undone but you don't know that"

may genuinely inform me. But we need not deny that learning involves acquiring knowledge in order to accommodate this possibility. Your statement is more accurately represented as

"Your shoelaces are *now* undone but you don't *now* know that."

Call the time at which you start making this statement, t_0, the time at which you finish stating its first conjunct, t_1 and the time at which you finish making the whole conjunctive statement, t_2. Assuming that you have just noticed that I am unaware of my undone shoelaces and wish to voice that thought immediately, your statement says, roughly,

"Your shoelaces are undone at t_0 but you don't know that at t_0."

By t_1 I learn, and so come to know, that my shoelaces are undone at t_0. But this does not falsify the second conjunct of your statement, which says only that I do not know at t_0 that my laces are undone at t_0. By t_2 I learn, and so come to know, that I do not know, at t_0, that my shoelaces are undone at t_0. Since I retain my first piece of knowledge while you finish speaking, at t_2 I know that my shoelaces are undone at t_0. Since my second piece of knowledge is factive, it is true, at t_2, that I do not know, at t_0, that my shoelaces are undone at t_0. This is no contradiction. In other words, what I learn, and so come to know, is that my shoelaces were (and still

are) undone and that I did not know this before you made your statement. But this remains true after you have made your statement. By contrast, if you tell me

"Your shoelaces are undone but you will *never* know this"

then it would be impossible for me to come to know this. But then again, it would be equally impossible for me to learn that it is true, for if it is, then I cannot know its first conjunct.

ENDNOTES

1. Against solutions that construe the announcement as self-referring in such a way that makes it a necessary falsehood.

2. He actually uses "judgment" in place of "belief."

3. Sainsbury deals with the version of the paradox in which surprise is taken as the absence of foreknowledge.

4. For present purposes I evade an objection that De Almeida might make. In (2001) he objects to my (1994) account of the absurdity of believing Moore's '*p* and I do not believe that *p*'—that your believing this falsifies it, so you are *irrational* in believing it. His objection is that you are not irrational in forming a self-falsifying belief unless you *recognize* that it is self-falsifying. But recognizing this is no mean feat, so failing to recognize this won't make you irrational (De Almeida, 2007). He would make the same objection to a similar diagnosis of the irrationality of the Moorish belief that (*p* and I am not justified in believing that *p*)—that your becoming justified in this belief falsifies it, so you are *irrational* in believing it. His objection would be that you are irrational in the Moorish belief only if you *recognize* that any conceivable justification you might have for it falsifies it. Fortunately I need not face this objection here. In what follows I need only argue that you cannot be *justified* in the Moorish belief.

5. I have taken the liberty of equating the case in which *s* is justified in believing that *p* with the case in which she reliably forms and sustains her belief that *p*. I see that this is open to objection, but I think that what I say about justification can just be said again in non-reliabilist terms. De Almeida (2001, 53–54) defends this principle. He uses it in conjunction with elements of Klein's (1985; 1986) defeasibility theory of epistemic justification to reach the same conclusion that it is impossible for anyone to be justified in believing the Moorish proposition. My argument is simpler, because it does not rely on Klein's theory.

6. Peter Klein (1981, 79) may be the first to endorse Justification-Distribution, which he calls 'The Conjunction Principle.' He reasserts it in Klein (2007).

7. This might be denied. Someone might even claim that Justification-Distribution is synthetic *a priori*. But at least it is arguable that it partly constitutes belief and that Goldman's Principle partly constitutes being justified in belief, which is enough to suit my purpose. Bergmann (1997) claims that *every* major epistemology should incorporate some such mechanism as Goldman's Principle.

8. I hesitate to attribute this objection to De Almeida, as note 4 makes clear.

9. I have just considered a version of the paradox that takes surprise as the absence of *being justified in believing* that the exam will be on a particular day. A slightly different version results from taking surprise as the absence of *justified belief* that the exam will be on a particular day. The difference is that you may be justified in believing that p even when, for some reason, you have not yet formed the belief that p. What I have said about the first version will hold of the second provided we accept that justified belief distributes over conjunction and that if you justifiably believe that you do not justifiably believe that p then you do not justifiably believe that p.

A causal theorist of justification might reject the former by claiming that it is possible that what makes you believe that (p and q) satisfies some appropriate reliability conditions but what makes you believe that p does not; you might believe that (p and q) on reliable testimony, but believe that p because of wishful thinking. Suppose that you join a dating agency but forget to tick any of the preferences beginning with "tall," "dark," and so on. Your wish to date someone tall makes you think that your date will be tall. Then the agency, which you know is reliable, tells you that your date will be tall and dark. You have a good reason to think that your date will be tall and dark but a bad reason to think that she will be dark. So you have a justified belief that she will be tall and dark but you do not have a justified belief that she will be tall.

I reply that once you accept the testimony you no longer believe that your date is tall on the basis of wishful thinking—at least not if you are justified in your beliefs. But couldn't your belief be causally overdetermined? Perhaps, but in such a case we should attribute you the justification you would attribute yourself, which will not be your hopes but the testimony—at least if you are justified in your beliefs.

10. For example, Ayer (1973, 125); Cargile (1967, 550); Kaplan and Montague (1960, 80); Margalit and Bar-Hillel (1983, 279 and 264); and Nerlich (1961, 503–504).

11. Sainsbury (1995, 93) makes this point, but offers no way of faulting the strong *reductio* (1995, 95–96).

12. I owe this objection to Zach Weber.

13. I owe this reply to Claudio De Almeida.

14. It might be objected that since my analysis turns upon denying that the student's justification for believing the announcement persists throughout the week, it does not solve Sorensen's non-temporal version of the paradox (Sorensen, 1986). Appendix I explains why this is not so.

15. This reading is strongly supported by Quine's remark (1953, 66):

> Thus suppose that a mathematician at work on the Fermat problem assumes temporarily, for the sake of exploring the consequences, that Fermat's proposition is true. He is not thereby assuming, even as a hypothesis for the sake of argument, that he knows Fermat's proposition to be true.

However Sorensen (1988, 263–264) gives a different reading of Quine:

> Quine points out that the *reductio ad absurdum* is of the supposition that *K knows the announcement*; the *reductio* is not of the announcement itself. To conclude that the announcement cannot be true because *K* cannot know it to be true is to commit a modal version of *argumentum ad ignorantiam*.

Against Sorensen, it is not hard to understand why someone might confuse supposing that something is true with supposing that she knows that is true. By contrast, it is hard to accept that anybody would commit the fallacy of arguing, "I cannot know that the announcement is true, so the announcement cannot be true." This doesn't seem a reasonable explanation of why people are often taken in by the weak *reductio*.

16. At one point Quine seems to suggest that *K* may not be taken as offering any kind of *reductio* (1953, 66):

> The tendency to be deceived by the puzzle is perhaps traceable to a wrong association of *K*'s argument with *reductio ad absurdum*. It is perhaps supposed that *K* is quite properly assuming fulfillment of the decree, for the space of his argument, in order to prove that the decree will not be fulfilled.

This is one of the most misleading passages of Quine's murky paper. He clearly has the weak *reductio* in mind in continuing

> This, if it were all, would be a good reductio *ad absurdum*; and it would entitle *K* to eliminate *(b)* and *(c)*, but not *(d)*.

If *K* (or the student) is not offering a *reductio* then what is he doing? He must either start with the supposition that the announcement is true (or that the decree will be fulfilled) or with the supposition that he knows that it is true (or will be fulfilled) or with the supposition that he is justified in believing that it is true (or will be fulfilled). This, together with his knowledge of what has happened as the week progresses, is all he has to work with.

17. Two anonymous referees, each for a different journal, both made this objection.

18. A version of this paper was presented at the Australasian Association of Philosophy Conference, New Zealand Division, at the University of Otago, 03–08 December 2005 as part of a project funded by Singapore Management University Research Centre. I am especially indebted to Mark D'Cruz. Referees for this journal provided useful comments. Claudio De Almeida's sharp criticisms forced me to develop the paper in many ways.

19. Gerbrandy actually uses the word "believe," but he treats "believe" and "know" interchangeably (1999, 26).

20. In recent correspondence, Gerbrandy confirms that this summary of his solution is accurate.

BIBLIOGRAPHY

Adams, F. 2005. "Tracking Theories of Knowledge." *Veritas* 50: 11–35.

Adams, F., and M. Clarke. 2005. "Resurrecting the Tracking Theories." *Australasian Journal of Philosophy* 83: 207–221.

Ayer, A. J. 1973. "On a Supposed Antinomy." *Mind* 82: 125–126.

Bergmann, M. 1997. "Internalism, Externalism and the No-Defeater Condition." *Synthese* 110: 399–417.

————. 2005. "Defeaters and Higher-Level Requirements." *The Philosophical Quarterly* 55: 419–436.

Binkley, R. 1968. "The Surprise Examination in Modal Logic." *Journal of Philosophy* 65: 127–135.

Cargile, J. 1967. "The Surprise Test Paradox." *Journal of Philosophy* 64: 550–563.

De Almeida, C. 2001. "What Moore's Paradox Is About." *Philosophy and Phenomenological Research* 62: 33–58.

————. 2007. "Moorean Absurdity: An Epistemological Analysis." In *Moore's Paradox: New Essays on Belief, Rationality and the First Person*, ed. Mitchell. S. Green and John N. Williams. Oxford: Oxford University Press.

Gerbrandy, J. 1999. *Bisimulations on Planet Kripke*. Amsterdam: ILLC Dissertation Series.

————. 2007. "The Surprise Examination in Dynamic Epistemic Logic." *Synthese* 155: 21–33.

Goldman, A. 1986. *Epistemology and Cognition*. Cambridge, Mass.: Harvard University Press.

Jongeling, B., and T. Koetsier. 1993. "A Reappraisal of the Hangman Paradox." *Philosophia* 22: 299–311.

Kaplan, D., and R. Montague. 1960. "A Paradox Regained." *Notre Dame Journal of Formal Logic* 1: 79–90.

Klein, P. 1981. *Certainty: A Refutation of Scepticism*. Minneapolis: University of Minneapolis Press.

————. 1985. "The Virtues of Inconsistency." *Monist* 68: 105–135.

————. 1986. "Immune Belief Systems." *Philosophical Topics* 14: 259–280.

————. 2007. "Useful False Beliefs." In *Epistemology: New Essays*, ed. Quentin Smith. Oxford, Oxford University Press.

Margalit, A., and M. Bar-Hillel. 1983. "Expecting the Unexpected." *Philosophia* 13: 263–288.

Moore, G. E. 1942. "A Reply to My Critics." In *The Philosophy of G. E. Moore*, ed. P. Schilpp. Evanston: Tudor, 535–677.

Nerlich, G. 1961. "Unexpected Examinations and Unprovable Statements." *Mind* 70: 503–513.

Nozick, R. 1981. *Philosophical Explanations*. Cambridge, Mass.: Harvard University Press.

Olin, D. 1983. "The Prediction Paradox Resolved." *Philosophical Studies* 44: 255–233.

————. 1984. "The Prediction Paradox: Resolving Recalcitrant Variations." *Australasian Journal of Philosophy* 64: 181–190.

————. 2003. *Paradox*. Montreal: McGill-Queen's University Press.

Quine, W. V. O. 1953. "On a So-Called Paradox." *Mind* 62: 65–67.

Sainsbury, M. 1995. *Paradoxes* (Second Edition). Cambridge: Cambridge University Press.

Sorensen, R. 1984. "Conditional Blindspots and the Knowledge Squeeze." *Australasian Journal of Philosophy* 62: 126–135.

————. 1986. "A Strengthened Prediction Paradox." *The Philosophical Quarterly* 36: 504–513.

————. 1988. *Blindspots*. Oxford: Oxford University Press.

Veltman, F. 1996. "Defaults in Update Semantics." *Journal of Philosophical Logic* 28, 221–261.

Williams, J. N. 1994. "Moorean Absurdity and the Intentional 'Structure' of Assertion." *Analysis* 54: 160–166.

Wright, C. 1995. "Truth in Ethics." *Ratio* 8: 209–226.

Wright, C., and A. Sudbury. 1977. "The Paradox of the Unexpected Examination." *Australasian Journal of Philosophy* 55: 41–58.