The Solution to the Surprise Exam Paradox

Ken Levy
Louisiana State University Law Center

Abstract

The Surprise Exam Paradox continues to perplex and torment despite the many solutions that have been offered. This paper proposes to end the intrigue once and for all by refuting one of the central pillars of the Surprise Exam Paradox, the "No Friday Argument," which concludes that an exam given on the last day of the testing period cannot be a surprise. This refutation consists of three arguments, all of which are borrowed from the literature: the "Unprojectible Announcement Argument," the "Wright & Sudbury Argument," and the "Epistemic Blindspot Argument." The reason that the Surprise Exam Paradox has persisted this long is not because any of these arguments is problematic. On the contrary, each of them is correct. The reason that it has persisted so long is because each argument is only part *of the solution. The correct solution requires all three of them to be combined together. Once they are, we may see* exactly *why the No Friday Argument fails and therefore why we have a solution to the Surprise Exam Paradox that should stick.*

1. Introduction

While many solutions have been offered to the Surprise Exam Paradox (SEP) over the past sixty years, it continues to perplex and torment. This paper attempts to end the intrigue once and for all.

Many, if not most, people who have thought about SEP will initially—if not persistently—accept the conclusion that a sur-

Ken Levy is transitioning from Climenko Fellow at Harvard Law School to Assistant Professor at Louisiana State University Law Center. He received a PhD in philosophy from Rutgers University in 1999 and a JD from Columbia Law School in 2002. His recent publications include "The Rationalist Solution to Gregory Kavka's Toxin Puzzle" (Pacific Philosophical Quarterly, 2009), "Killing, Letting Die, and the Case for Mildly Punishing Bad Samaritanism" (Georgia Law Review, forthcoming 2010), "The Solution to the Real Blackmail Paradox: The Common Link Between Blackmail and Other Criminal Threats" (Connecticut Law Review), and "Baumann on the Monty Hall Problem and Single-Case Probabilities" (Synthese).

prise exam cannot be given on the very last day of the testing period. For the argument that an exam on the last day of the testing period will *not* be a surprise—what I will call the "No Friday Argument"—seems irrefutable. After all, if a surprise exam has been promised, especially by a trustworthy teacher, and there is now only one day left in the testing period, then it *must* be given on that last day. There just does not seem to be any other possible conclusion.

But there is. And the task of solving—really *solving*—SEP requires it to be shown, clearly and carefully, just how an exam on this last day still *would* be a surprise. The first step in accomplishing this task is to expose the flaw in our reasons for accepting the No Friday Argument. The second step is to demonstrate the flaw in the No Friday Argument itself.

There are two main intuitions that drive us initially to accept the No Friday Argument. First, we have the intuition that the student cannot be surprised by an exam that she *expected* the night before. Second, we have the intuition that a student cannot be surprised by an exam that she was *certain* the night before would take place the next day. Given both of these intuitions and three facts—the fact that an ex hypothesi trustworthy teacher promised an exam by the end of the testing period, the fact that an exam has not yet been given, and the fact that there is now only one day left in the testing period—it seems indisputable that the student will not be surprised by an exam on the last day of the testing period.

Upon reflection, however, this conclusion is both doubtful and false. First, it is doubtful because the two intuitions behind it are doubtful. The two intuitions behind it are doubtful because *surprise* in the context of SEP must be interpreted as the absence of *knowledge*—that is, the absence of a justified-and-true belief—the night before that the exam will take place the next day. And neither mere expectation nor (psychological) certainty is sufficient for knowledge.

Second, the conclusion that an exam on the last day will not surprise the student is *false*. There are three arguments in the literature that establish this point. Once they are put together, the reader will be able to see that, and exactly why, the No Friday Argument is wrong. And once the No Friday Argument has been thoroughly refuted in this way, SEP has been (thoroughly) solved. This paper, then, aspires to present the definitive solution to SEP.

The first of these three anti–No Friday Arguments is the "Unprojectible Announcement Argument." The Unprojectible Announcement Argument suggests that the No Friday Argument begins with a mistaken assumption: that the student may unproblematically project the teacher's announcement of a surprise exam forward to the evening before the last day of the testing period. In fact, the student must *change* her interpre-

tation of the announcement from "There will be a surprise exam by the end of the testing period" to "There will be a surprise exam *tomorrow*." And this statement either is, or borders on being, self-contradictory. Either way, the student is not, at the time of the teacher's announcement of a surprise exam (not to mention afterward), epistemically entitled to assume in the first place that the teacher's announcement is just as applicable on the night before the last day of the testing period as it was when the teacher first made the announcement.

Second, the "Wright & Sudbury Argument" is a somewhat simpler version of Crispin Wright and Aidan Sudbury's argument that while the student starts the testing period with the knowledge that the surprise exam will occur during the testing period, she will lose this knowledge by the evening before the last day of the testing period (if the exam still has not been given by then), in which case an exam on the last day will still be a surprise.

The Wright & Sudbury Argument, however, cannot be the end of the story. For the conclusion of the Wright & Sudbury Argument—again, that an exam on the last day will be a surprise—is vulnerable to a reductio, what I will call the "Renewed No Friday Argument." But I will argue that *this* argument cannot be the end of the story either, for *its* conclusion is equally vulnerable to an equal and opposite reductio, once again the Wright & Sudbury Argument. So we end up with a logical circle much like that generated by the Liar Paradox statement, "This statement is false." This circle is arguably the heart of SEP, the reason that SEP continues to baffle and solutions continue to be proposed.

I will argue that what resolves this problem, what "dissolves" this logical circle, is yet a third argument, the "Epistemic Blindspot Argument." The endless logical circle generated by the Wright & Sudbury Argument and the Renewed No Friday Argument is succinctly captured by Roy Sorensen's concept of an "epistemic blindspot," the notion that the student is in a situation such that if the exam has not occurred by the penultimate day of the testing period, then the student will justifiably believe that the exam will be given the next day if and only if it will *not*. So whatever the student justifiably believes about an exam on the last day of the testing period helps to make the opposite the case. If she justifiably believes that it will be a surprise, then it will not be; conversely, if she justifiably believes that the exam will *not* be a surprise, then it *will* be. The Epistemic Blindspot Argument then continues: because the student has no good reason to adopt one of these beliefs over the other, she is not justified in believing either, in which case she is *not* justified in believing that the exam will be on the last day. So, once again, contrary to the No Friday Argument, an exam on the last day *will* be a surprise.

2. The Impossibility Argument

In this part, I will explicate the assumptions that I take to be essential to the strongest version of SEP.

2.1 Clever

Suppose that a particular teacher has a class that meets on Tuesday through Friday and consists of only one "ideally" or "optimally" or "maximally" rational student—let's call her "Clever."[1] By *maximally rational*, I mean "a master logician; [s]he avoids contradiction, is aware of all logical truth, and believes all the logical consequences of what [s]he believes." Moreover, she always knows who "[s]he is and on what occasion [s]he is judging ... [and] which occasions are earlier and later than which" (Binkley 1968, 129).[2] (I will defend the maximal rationality assumption in section 2.6 below.)

2.2 The Announcement

Suppose also that the teacher tells Clever on Tuesday that he will be giving her an exam this week on either Wednesday, Thursday, or Friday (exclusive) and that it will be a surprise in the sense that it will occur on a day that follows a night on which Clever could not have known that the exam will be the next day.[3] Call this the "Announcement" that the teacher gives to Clever. For the sake of convenience, I will refer to the Announcement as "(A)", to the part of (A) that promises an exam this week as "$(A)_E$", and to the part of (A) that promises that this exam will be a surprise as "$(A)_S$".

2.3 The No Friday Argument

After the teacher issues (A) to Clever, Clever gives him the "Impossibility Argument." The Impossibility Argument attempts to show that (A) is self-defeating. It consists of two subarguments, the "No Friday Argument" and the "Backtracking Argument." The No Friday Argument says that if the exam is not given by Thursday, then Clever will know on Thursday night that the exam will be on Friday. So if the exam is given on Friday, the teacher's promise that the exam will be a surprise will be violated. Given (A), then, Friday is not really a possible day for the exam.

2.4 The Backtracking Argument

The Backtracking Argument picks up where the No Friday Argument leaves off. The fact that a surprise exam cannot be given on Friday means that if the teacher has not given the exam on

Wednesday, Clever will know on Wednesday night that the exam will be on Thursday. For she already knows on Wednesday night that Friday is no longer a possible day for the exam. So an exam on Thursday would also violate the teacher's promise that the exam will be a surprise. Given (A), then, Thursday—like Friday—is no longer a possible day for the exam either. Since Thursday and Friday are no longer possible days for the exam, it follows that the teacher must give the exam on Wednesday. But then Clever can figure this out on Tuesday night, in which case an exam on Wednesday would not be a surprise either. Therefore the teacher simply cannot give a surprise exam. Despite his best intentions, he simply cannot fulfill (A).[4]

2.5 The Intuition

The Impossibility Argument suggests that when we reason backwards from Friday to Wednesday, it turns out that a surprise exam is impossible. But this conclusion cannot be the end of the story. For when we consider "moving forward" from Wednesday through Friday, a surprise exam suddenly seems possible again. We have the intuition—call it "the Intuition"—that if the teacher gave the exam on at least Wednesday or Thursday, the exam would indeed surprise Clever in the sense that she would not have known the night before that the exam would be the next day rather than on one of the remaining days. So something must be wrong. A surprise exam cannot be both possible and impossible. Likewise, one and the same exam cannot be both a surprise and not a surprise. Instead, one of these two approaches—either the Impossibility Argument or the Intuition—must be wrong. The challenge of SEP, then, is to determine which of these it is and why.[5]

2.6 The Maximal Rationality Assumption

I assumed in section 2.1 above that Clever must be maximally rational. The reason is that, in order to solve a paradox, we need to consider its strongest version. A solution to any weaker version of the paradox still leaves the stronger version(s) unsolved (see Jongeling and Koetsier 1993, 300, and Olin 1983, 226).[6] As it happens, the more rational Clever is, the stronger SEP is. For the more rational Clever is, the more trivial solutions to SEP we may rule out—for example, Clever forgets the teacher's announcement, does not fully understand (A), or makes invalid inferences. So only if Clever is maximally rational may we rest assured that SEP is maximally strong (see Ferguson 1991, 300; Hall 1999, 650–51; Olin 1983, 227; 1988, 114; and Sorensen 1984, 130). Still, we must be careful to distinguish maximal rationality from *infallibility*. For one of the central points of this paper will be that Clever's Impossibility Argument

fails. It will be argued, however, that the Impossibility Argument fails *not* because it is invalid, in which case we *would* have to conclude that Clever is *not* maximally rational, but rather because Clever adopts premises that, while perfectly plausible on their face, turn out on closer inspection to be false. The latter scenario is possible even if Clever is maximally rational. Clever's being maximally rational protects her only from making silly mistakes and invalid inferences, not from starting with initially plausible but provably false premises.

2.7 Knowledge and Surprise

I have framed (A) and the Impossibility Argument in terms of the word *know*. I have said that, according to (A), an exam will be a surprise if and only if it occurs on a day that follows a night on which Clever could not have *known* that the exam would be the next day; and that, according to the Impossibility Argument, Clever can *know* on any night that the exam will be the next day. The reader should understand that knowledge here is assumed to be justified-and-true belief.[7] To be sure, the Gettier problem suggests that this definition of knowledge is inadequate. But we do not need to consider the specific respects in which it is inadequate—and therefore the specific refinements required to make the definition "Gettier-proof"—for the purposes of SEP (see Williams 2007, 78–79).

Given that surprise is the absence of knowledge and that knowledge is justified-and-true belief, an exam is a surprise if and only if Clever did not justifiably believe the night before that it would now take place.[8] This means that (A) is equivalent to the following disjunctive proposition: (a) there will be an exam Friday, and Clever will not be justified in believing this on Thursday night; or (b) there will be an exam Thursday, and Clever will not be justified in believing this on Wednesday night; or (c) there will be an exam Wednesday, and Clever will not be justified in believing this on Tuesday night. We need not worry about situations in which Clever justifiably but *falsely* believes that an exam will take place the next day. She will *not* be surprised in the relevant sense the next day because, as we have already seen in section 2.2 and note 3 above, Clever may be surprised only by an exam, not by the nonoccurrence of an exam.

Some might argue that knowledge is too strong for SEP; that SEP may be formulated instead in terms of mere *belief*, whether justified or not (see Cave 2004, 610; Goldstein 1993; Rescher 2001, 112–14; and Schick 2000). But this "mere belief" version of SEP would render (A)$_S$ nonsensical and the Impossibility Argument unworkable. There are two reasons. First, (A)$_S$ would make little sense on the belief version because it would involve the teacher's making a promise that is nonsensical on its face—that is, a promise to give an exam that will occur on a day that

follows a night on which Clever does not *believe*, justifiably *or unjustifiably*, that the exam will be given the next day (see Kiefer and Ellison 1965, 426). Second, the contradiction between the Impossibility Argument and the Intuition is what renders SEP a *paradox*. But if *surprise* meant prior nonbelief as opposed to prior non–*justified-and-true* belief, the Intuition—and therefore the paradox—would disappear. For we do *not* have an intuition that Clever cannot *un*justifiably believe on Tuesday night that the exam will be on Wednesday or on Wednesday night that the exam will be on Thursday. We think that this situation is perfectly possible. So if *surprise* were defined as prior nonbelief, we would be *perfectly willing* to accept the Impossibility Argument (see Kiefer and Ellison 1965, 426).

2.8 A Failed, but Instructive, Argument against the Backtracking Argument

One might attempt to refute the Backtracking Argument as follows. To say that Friday is not a possible day for the exam is clearly false. Regardless of what the No Friday Argument proves, it is still perfectly *physically* possible for the teacher to give the exam on Friday. It is just that if the teacher *did* give it then, he would not be living up to his promise to give a *surprise* exam. So it is not that Friday is not a possible day for the exam in the sense that Saturday is not a possible day for the exam. Rather, a better way to put the conclusion of the No Friday Argument is that Friday is not a possible day for the exam *if* the teacher is to fulfill (A). But once we put the conclusion of the No Friday Argument in this conditional form, then it seems that Clever may no longer simply assume that Friday is not a possible day for the exam and thereby limit on Tuesday night the possible days of the exam to Wednesday and Thursday. For Clever cannot simply assume that the antecedent (i.e., the teacher is to fulfill (A)) is true.

The objection continues: as long as giving the exam on Friday remains a *physical* possibility, Clever may not justifiably believe on Wednesday night that the exam will be given Thursday rather than Friday (see Janaway 1989, 402). Moreover, whatever misgivings Clever may have with regard to her ruling out Friday on Wednesday night as a possible day for the exam are only compounded on Tuesday night. Since Clever's ruling out Thursday presupposes her ruling out Friday, and since she already has some misgivings about ruling out Friday, she will have even more misgivings about ruling out Thursday as well. So as the number of possible days for the exam gets larger and larger, Clever's—and therefore our own—misgivings about the Backtracking Argument will tend to get correspondingly larger and larger as well (see Hall 1999, 650, 652, 660–61, 682–83; Janaway 1989, 404; and Williamson 2000, 142).

137

A supporter of the Impossibility Argument will respond to this objection as follows. Yes, it always remains *physically* possible for the teacher to give the exam on Friday, no less on Thursday or Wednesday. But this point is not really relevant. The No Friday Argument shows that it is not possible—not *logically* possible, no less *physically* possible—for the teacher to give the exam on Friday *and* fulfill (A) (see McLelland 1971, 82). And Clever *must* assume that the teacher will fulfill (A) if he can (see Hall 1999, 653ff.).[9] (I will defend this assumption further in the next section.) So as long as Clever hangs on to this assumption that (A) will be fulfilled if it can be fulfilled, she *can* conclusively rule out Friday. She should have absolutely no misgivings about this assumption on Wednesday night. And if Clever should not have any misgivings about ruling out on Wednesday night a Friday exam, then she should not have any compounded misgivings about ruling out on Tuesday night a Thursday exam. In other words, once Clever recognizes that the teacher will fulfill (A) if he can, she no longer has any good reason to abandon the Backtracking Argument. Or so the proponent of the Impossibility Argument argues.

2.9 The Teacher Will Fulfill (A) If He Can

In this section, I will defend my statement in the previous section that Clever—and therefore we—must assume that the teacher will fulfill (A) if he can.

There are two reasons. First, we just saw in the previous section that a supporter of the Impossibility Argument clearly relies on this premise. So it begs the question against the Impossibility Argument to assume that the teacher may *not* fulfill (A) even if he can. Of course, if this premise—that is, the premise that the teacher will fulfill (A) if he can—were obviously erroneous, then we would have to reject it. But it is not obviously erroneous. So without an argument for this conclusion, we may not simply dismiss this premise, at least not if we are going to be fair to Clever.

Second, to reject this assumption would be to violate one of the key stipulations that I made in section 2.6—namely, that we need to consider the strongest version of SEP. A version of SEP according to which Clever assumes that the teacher will fulfill (A) if he can is stronger than a version of SEP according to which Clever abandons this assumption. For were Clever to abandon this assumption, the Impossibility Argument—in particular, the Backtracking Argument—would be much too easy to refute. We could simply reason as follows:

(1) Assume that the teacher will *not* necessarily fulfill (A) if he can.

(2) ∴ Clever is not justified in believing (A). [(1)]

138

(3)   The only basis that Clever has for believing either part of (A)—that is, $(A)_E$ or $(A)_S$—is (A) itself. Clever has no *independent* reason—no reason *other than* (A)—to believe $(A)_E$ or $(A)_S$ (see Hall 1999, 649, 653, 654).[10]

(4)   ∴ If Clever is not justified in believing (A), then she is not justified in believing either part of (A). [(3)]

(5)   ∴ Clever is not justified in believing $(A)_E$. [(2), (4)]

(6)   ∴ Clever is not justified in believing on any given night that there will be an exam the next day—even if no exam has been given by Thursday and only Friday is left. [(5)]

(7)   ∴ An exam on any given day—even Friday—would be a surprise. [(6)]

To be sure, this reasoning *does* solve one version of SEP—namely, a version according to which (1) is the case. But it does not solve the version according to which (1) is *false*. So our assuming (1) hardly ends our task. We still need to determine whether SEP can be solved if we adopt the opposite assumption (see Cave 2004, 612). Indeed, we might as well assume that the teacher is an infallible computer.

It is important to recognize and acknowledge that the argument above, (1)–(7), moves in a rather bizarre direction. While it starts from the assumption that the teacher will not necessarily fulfill (A), it arrives at the virtually opposite conclusion that an exam on any day *would* fulfill (A), in which case the teacher *will* fulfill (A) as long as he gives an exam. This ironic twist, which foreshadows the logical circle that will be discussed in section 5.3 and the "epistemic blindspot" in section 5.4, is ultimately generated by the tension between the teacher's announcement of (A), which justifies Clever's belief that an exam will take place, and $(A)_S$, which *combats* justification of Clever's belief on any given night that the exam will take place the next day. As we will see below in section 5.1, this tension becomes especially acute if an exam has not been given by Thursday and there is therefore only one day left in the testing period.

3. The Impossibility Argument Cannot Be Correct

If the Impossibility Argument is correct, then a surprise exam cannot be given on any day. And to say that a surprise exam cannot be given on any day is to say that (a) on Tuesday night, Clever is justified in believing that there will be an exam on

Wednesday; (b) if Wednesday passes without an exam, then Clever is justified in believing on Wednesday night that there will be an exam on Thursday; and (c) if Thursday passes without an exam, then Clever is justified in believing on Thursday night that there will be an exam on Friday (see Austin 1969).

In order to see why the Impossibility Argument cannot be correct, it might help to expand the number of days in the testing period. For as I suggested in section 2.8, the greater the number of days in the testing period, the stronger the Intuition. And the stronger the Intuition, the higher the hurdle that the Impossibility Argument must jump. So let's assume two things. First, let's assume now that there are not three but three *hundred* days in the testing period. Second, let's assume that the exam is given on, say, the 155th day. By the Impossibility Argument, then, Clever was not surprised by this exam. For, again, the Impossibility Argument states that a surprise on *any* day of the testing period cannot be a surprise. And to say that an exam on the 155th day was not a surprise is just to say that Clever had a justified belief on the 154th night that the exam would be given the next day. Again, this conclusion is entailed by the Impossibility Argument (in conjunction with the two other assumptions—an expansion of the testing period to three hundred days and an exam on the 155th day).

Given this expansion of the testing period, it is easier to understand why the Impossibility Argument cannot be correct. If (ex hypothesi) the exam occurs on the 155th day, then Clever is merely lucky that her belief the night before that the exam would be the next day turned out to be true. She is merely lucky because her belief turned out to be false all other nights. And she was in no better epistemic position on this particular night than she was on any of the other previous 153 nights. She had no more reason than she did on any previous night to believe that the exam would be given the next day. Therefore her belief on the 154th night—like her belief on the first 153 nights—was really not justified to begin with (see Cargile 1967, 552; Gardner 1963, 152; Halpern and Moses 1986, 284; and Williamson 1992, 223–24).[11]

Since the three-hundred-day scenario differs only in degree and not in kind—that is, does not require a different kind of solution—from the three-day scenario,[12] what applies to the former applies to the latter. If it is not the case that Clever has a justified belief every night that the exam will be given the next day in the three-hundred-day scenario, then it is not the case in the three-day scenario either. So the Impossibility Argument fails.[13]

4. The Misleading Intuitions Argument

The No Friday Argument is quite powerful. As I will explain in this section, however, its persuasiveness draws from two intui-

tions of ours. And these intuitions mislead us. They lead us initially to accept the conclusion of the No Friday Argument when it has not actually been established. Call this the "Misleading Intuitions Argument." The Misleading Intuitions Argument is designed to show *not* the stronger point that the No Friday Argument is fallacious—that will be the task of section 5—but rather the slightly weaker point that we may initially be inclined to accept it for what turn out to be bad reasons.

Suppose that it is Thursday night and the teacher has not yet given the exam. Our first intuition is that the exam will be on Friday. Once again, the teacher, an ex hypothesi trustworthy individual, if not infallible computer, said that he would give an exam by the end of the testing period, and here it is—the end of the testing period. These facts are initially sufficient to motivate our—and therefore Clever's—belief that an exam will be the next day. The No Friday Argument then easily concludes that an exam on Friday will not be a surprise. After all, if Clever is expecting an exam, then it cannot be a surprise.

What this version of the No Friday Argument subtly overlooks is that *more* than this expectation is necessary to show that an exam on Friday will not be a surprise. It also needs this expectation to be *justified*. As we learned in section 2.7, surprise is to be understood in terms of knowledge, which is itself to be understood in terms of justified-and-true belief. Surprise cannot be understood merely in terms of belief/expectation alone, justified or unjustified. For, again, if it were, then we would no longer have the Intuition (that a surprise exam can be given on most, if not all, days of the testing period), in which case SEP would no longer be a paradox.

Even with this clarification, however, the No Friday Argument still seduces. Our second intuition is not merely that Clever can expect on Thursday night an exam on Friday but that Clever can be nearly *certain* about it. And, again, this conclusion tempts us to accept the No Friday Argument and move on to the Backtracking Argument. But even if Clever could be certain on Thursday night that the exam would be on Friday, the No Friday Argument might still fail. For certainty is not necessarily sufficient for justified belief. The degree to which Clever's beliefs are justified does not necessarily bear any correlation with the degree to which she is certain of them. Just as she may confidently hold unjustified beliefs, she may inconfidently hold justified beliefs. "Justified-ness" is a normative property of beliefs; certainty or confidence, a psychological property or attitude. (Importantly, I have in mind *psychological* certainty, not *propositional* certainty. The former is the belief that a certain proposition—*p*—is true and that there are no legitimate grounds whatsoever to doubt that *p* is true; the latter is *full warrant* in this belief [see Klein 1998].) So they may very well come apart. Therefore even if the exam were not given by Thursday, Clever

might still confidently maintain an *un*justified belief on Thursday night that the exam will be on Friday. If so, then given the definition of *surprise*, such an exam would still be a surprise.

5. Three Arguments against the No Friday Argument

Now that we have exposed the two powerful, but misleading, intuitions that underlie our initial adherence to the No Friday Argument, it should be easier to accept some arguments against it, arguments that help to show where it goes wrong.

5.1 Argument 1: The Unprojectible Announcement Argument (and Moore's Paradox)

The first argument against the No Friday Argument—the "Unprojectible Announcement Argument"—starts with the observation that the No Friday Argument rests on the implicit assumption that the teacher's Announcement—(A)—may be applied to Thursday night, the last night of the testing period. Again, when the teacher first issues (A), he promises Clever a surprise exam by the end of the testing period. The No Friday Argument then directly applies (A) to Thursday night and concludes that a surprise exam cannot be given on Friday.

What easily goes unnoticed is that this application of (A) to Thursday night assumes a significant shift in interpretation. It assumes that Clever's interpretation of (A) shifts from "There will be a surprise exam by the end of the testing period" to "There will be a surprise exam *tomorrow*" (see Wright and Sudbury 1977, 50, 54). Call this latter interpretation of the announcement "(A*)". Like (A), (A*) is composed of two propositions—$(A^*)_E$ and $(A^*)_S$. $(A^*)_E$ is the statement that Clever will have an exam on Friday; $(A^*)_S$ is the statement that Clever is not justified in believing on Thursday night that she will have an exam on Friday.

This conjunction of propositions is strange—so strange that we should question the very applicability of (A) to Thursday night in the first place. $(A^*)_E$ and $(A^*)_S$ together are (so) strange because they seem to create an untenable epistemological situation for Clever. On the one hand, $(A^*)_E$ gives Clever justification for believing on Thursday night that an exam will be given on Friday. That is, the teacher's stating to Clever that she will have an exam the next day is arguably equivalent to the teacher's stating that he, the teacher, is *authorizing* Clever to believe—and therefore *justifying* her in believing—that she will have an exam the next day. On the other hand, $(A^*)_S$ states that Clever is *not* justified in believing that an exam will be given the next day. (A*), then, seems to be presenting Clever with a flat-out contradiction.[14]

One might try to defuse this contradiction by considering Claudio de Almeida's work on Moore's Paradox (see de Almeida 2001 and 2007).[15] Consider the following statement:

(8) It is raining and I do not believe it.

This statement certainly seems strange. The problem of Moore's Paradox, traditionally conceived, is to explain *why* (8) is strange, what the nature (or cause) of this "strangeness" is. But de Almeida conceives the problem of Moore's Paradox to extend not merely to statements exemplified by (8)—namely, statements of the type

(9) *p* and I do not believe *p* (or *p* and I believe ~*p*)

but also to statements involving *justified* belief. In this section, we are considering just such a statement—namely, (A*):

(10) There will be an exam tomorrow and I (Clever) am not justified in believing that there will be an exam tomorrow.

Statement (10) instantiates statements of the more general type:

(11) *p* and I do not justifiably believe *p* (or *p* and I justifiably believe ~*p*) (see Williams 2007, 73).[16]

In two different papers, de Almeida (2001, 2007) attempts to provide a general, unified explanation of why both (9) and (11) are "Moore-absurd" and therefore irrational for me, the asserter of the propositions, to believe.

de Almeida's earlier explanation differs somewhat from his later explanation. While his earlier explanation is "evidentialist," his later explanation is "non-evidentialist" (see de Almeida 2007, 70). What both explanations share in common, however, is an emphasis on the point that Moore-absurdity does *not* derive from self-contradiction (see de Almeida 2001, 38–43; 2007, 59–60). Moore-absurd statements cannot be self-contradictory because they may very well be true. It may very well be the case that *p* and I, the asserter of (9), do not believe *p*. Likewise, it may very well be the case that *p* and I, the asserter of (11), do not justifiably believe *p*. Instead, according to de Almeida, Moore-absurdity derives from the fact that (9) and (11) are "contradiction-like" (see de Almeida 2001, 34; 2007, 64). In (9) and (11), in addition to asserting that I do not (justifiably) believe *p*, I am *also asserting p* itself. And while my assertion of *p* does not *contradict* my assertion of my not (justifiably) believing *p*, it does *resemble* or *approximate* contradiction. The task is then to explain this "contradiction-likeness"—that is, why (9) and (11)

143

Ken Levy

initially *appear* to be self-contradictory, why they are not *in fact* self-contradictory, and why it is still irrational to believe them.

In his earlier paper on Moore's Paradox, de Almeida (2001, 51–52) argues that what makes my belief in (9) and (11) Moore-absurd and therefore irrational is that my belief in each proposition itself supplies evidence against—or reason to disbelieve—one of each proposition's two conjuncts. Consider first (9)—again, *p* and I do not believe *p*. My belief in (9) distributes over both conjuncts. So I believe each of the conjuncts—that *p* and that I do not believe *p*. Whether through introspection or from its mere presence in my belief system, my belief that *p* gives me reason to *believe that* I believe *p*. And this reason to believe that I believe *p* is evidence against the latter conjunct (that I do not believe *p*). My belief in (9) works against itself. It is therefore *self-refuting* or "epistemically self-defeating" (though, again, not self-contradictory). It gives me reason to disbelieve one of its own two conjuncts. And it is its providing evidence against itself that makes it irrational to believe.

de Almeida's analysis of (11)—that is, *p* and I do not justifiably believe *p*—is similar to, though not identical with, his analysis of (9) (see de Almeida 2001, 52–53). For this time, rather than arguing that the former conjunct "blocks" the latter conjunct, de Almeida argues just the reverse—that the *latter* conjunct blocks the *former* conjunct. To say that I do not justifiably believe *p* is to say that I do not have good reason to believe *p*. And my not having good reason to believe *p* itself constitutes a reason to *drop p* from my belief system. Yet my belief that (11) entails, through distribution, that *p* is *in* my belief system —that is, that I *do* believe *p*. So, once again, I have an epistemically self-defeating belief. My belief in (11) entails, through distribution, my not justifiably believing *p*; my not justifiably believing *p* constitutes reason for me to disbelieve *p*; my reason for disbelieving *p* weighs against the first conjunct—that is, *p*; and my reason to disbelieve the first conjunct constitutes reason to disbelieve the conjunction—(11), the original starting point of this "belief chain"—itself.

de Almeida's later explanation of Moore's Paradox goes down a different track (see de Almeida 2007, 64–69). According to de Almeida, his later explanation has the advantage over the earlier of being available not only to internalist-foundationalists (as the earlier is) but also to externalist-foundationalists (which the earlier is not) (see de Almeida 2007, 70). de Almeida's later explanation starts with this observation: the mere fact that statements like (9) (*p* and I do not believe *p*) and (11) (*p* and I do not justifiably believe *p*) are self-inconsistent—that is, contain inconsistent belief-conjuncts—does not by itself explain why they are incoherent with my belief system and therefore irrational for me to believe. For it is *rational* to believe *some* self-inconsistent propositions—namely, non–truth-functional

144

necessary falsehoods. Instead, according to de Almeida, Moore-absurd statements produce incoherence and irrationality not merely because they are self-inconsistent but because they are *strongly* inconsistent. A belief, *q,* is *strongly inconsistent* with my belief system—and is therefore irrational—if (a) there is a set of propositions in my belief system that entails ~*q* or (b) *q* is a truth-functional falsehood. Such is the case with both (9) and (11). My belief in (9) is strongly inconsistent with my not believing *p* because my belief in (9) entails a belief that *p,* and my belief that *p* conflicts with my not believing *p.* And my belief in (11) is strongly inconsistent with my justifiably believing ~*p* because (11) entails my believing *p*; and my believing *p* conflicts with my *justifiably* believing its opposite—that is, ~*p.*[17]

If we accept either of de Almeida's proposed explanations of Moore-absurdity, then we need not assume that (11) above—or (A*)—presents Clever with a *contradiction.* We may assume instead that it presents her "merely" with strong inconsistency. But this strong inconsistency is still sufficient to call into question the applicability of (A) on Thursday night—an application that, once again, turns (A) into (A*). For just as it is irrational to have self-contradictory beliefs, so too it is irrational to have strongly inconsistent beliefs (see also Williams 2007, 73–75).

Indeed, it is this worry about irrationality that has motivated the so-called Pragmatic Solution to SEP.[18] According to the Pragmatic Solution, in order to minimize the irrationality that (A)'s application to Thursday night engenders, it is more appropriate to interpret (A) on Thursday night not as (A*) but rather as a pragmatic statement—something like, "There will be an exam tomorrow, and it will not—as initially promised—be a surprise. Though unfulfilled, (A)$_S$ hopefully served its *real* purpose, which was to get you to study every night." This pragatic interpretation arguably makes better sense of (A) than does (A*) on Thursday night. But there is a high price to be paid for this gain in sensibility. The pragmatic interpretation of (A) arguably begs the question against the No Friday Argument and therefore the Impossibility Argument. For, again, the No Friday Argument *depends* on the implicit assumption that we may interpret (A) as (A*) on Thursday night.

5.2 Argument 2: The Wright & Sudbury Argument

The second argument against the No Friday Argument starts with the proposition from section 2.9: the teacher will fulfill (A) if he can. But *can* he any longer fulfill (A) as of Thursday night? Can the teacher still give an exam on Friday that will be a surprise? As we have seen, the No Friday Argument says no. But we may construct a powerful reductio ad absurdum of the No Friday Argument—call it the "Wright & Sudbury Argument" (see Wright and Sudbury 1977). The central thesis of the Wright

& Sudbury Argument is that while Clever may have been justi-
fied in believing (A) when the teacher first announced it, she
will *lose* this justification if the exam has not been given by the
penultimate day of the testing period:[19]

(12)   Assume that the No Friday Argument is correct. An
        exam on Friday would not be a surprise.

(13)   ∴ If the exam has not been given by Thursday, it is no
        longer possible for the teacher to fulfill $(A)_S$. [(12)]

(14)   ∴ Clever is no longer justified in believing $(A)_S$.[20] [(13)]

(15) $(A)_S$ is an essential part of (A).

(16)   ∴ Clever is no longer justified in believing (A). [(14),
        (15)]

(17)   If Clever is no longer justified in believing (A), then
        she is no longer justified in believing either part of
        (A). [See (4) in section 2.9. Once again, the teacher's
        announcement of (A) is the *only* reason that Clever
        has for believing either conjunct in the first place.]

(18)   ∴ Clever is no longer justified in believing $(A)_E$.[21] [(16),
        (17)]

(19)   ∴ Clever is no longer justified in believing on Thursday
        night that an exam will be given on Friday. [(18)]

(20)   ∴ An exam on Friday will be a surprise. [(19)]

(21)   Contradiction. [(12), (20)]

(22)   ∴ (12) is false. An exam on Friday *will* be a surprise.
        [(12), (21)]

5.3 The Renewed No Friday Argument, a Logical Circle, and a Resemblance to the Liar Paradox

The Wright & Sudbury Argument is strong, but it is hardly the
end of the matter. As it happens, from the conclusion of the
Wright & Sudbury Argument—that is, a Friday exam will be a
surprise—we may construct a compelling argument for the very
opposite conclusion, the conclusion of the No Friday Argument.
Call it the "Renewed No Friday Argument":

(23)   Assume that the conclusion of the Wright & Sudbury
        Argument is correct: an exam on Friday will be a
        surprise.

(24)   The teacher will fulfill (A) if he can. (See section 2.9.)

(25) ∴ Clever is justified in believing on Thursday night that an exam will be given on Friday. [(23), (24)]

(26) ∴ An exam on Friday will not be a surprise. [(25)]

(27) Contradiction. [(23), (26)]

(28) ∴ (23) is false. An exam on Friday will *not* be a surprise. [(23), (27)]

The Renewed No Friday Argument is persuasive, but it would be foolish to conclude at this point that the Wright & Sudbury Argument has been refuted. For we are faced with a rather strange situation. On the one hand, the Renewed No Friday Argument constitutes a reductio of the conclusion of the Wright & Sudbury Argument. On the other hand, the Wright & Sudbury Argument constitutes a reductio of the conclusion of the Renewed No Friday Argument. More concretely, as soon as Clever comes to believe that she cannot be surprised, she has put herself in a position where she *can* be surprised; and, conversely, as soon as Clever comes to believe that she *can* be surprised, she has put herself in a position where she *cannot* be surprised. We cannot stop at the conclusion of either argument because each conclusion serves as the first premise for the other argument. As soon as one argument ends, the other argument for the directly opposite conclusion has already begun. We seem, then, to be inescapably caught in an endless circle of reasoning, oscillating between two contradictory propositions (see Cave 2004, 609–13).

One might argue that this circle is unrecognized—unheard of—in analytic philosophy and therefore that one of the two arguments must be flawed. I agree that this circle is troublesome—arguably the heart of SEP. But it certainly is not unheard of. Indeed, the reader may notice that the circle that I have described here is similar to the logical circle in the Liar Paradox. In one version of the Liar Paradox, we are confronted with the following proposition:

(29) This statement is false.

Proposition (29) runs into a circle very much like the circle presented by the combination of the Wright & Sudbury Argument and the Renewed No Friday Argument. Just as the Wright & Sudbury Argument starts with the assumption that an exam on Friday will not be a surprise and ends with the conclusion that an exam on Friday will be a surprise, the assumption that (29) is false (immediately) leads to the conclusion that (29) is true. And just as the Renewed No Friday Argument starts with the assumption that an exam on Friday will be a surprise and ends with the conclusion that an exam on Friday will not be a surprise, the assumption that (29) is true (immediately) leads to the conclusion that (29) is false.

Perhaps both SEP and the Liar Paradox force us to acknowledge the possibility of a third kind of circle in addition to vicious circles (question-begging arguments) and virtuous circles (e.g., inference to the best explanation).[22] Still, whatever the merit of this suggestion, it would be hasty to infer from the fact that both SEP and the Liar Paradox generate endless circles of reasoning either (a) that the former reduces to a version of the latter or (b) that they are both susceptible to the same kind of solution. Regarding (a), while the two problems may initially lead to endless circles of reasoning, their sources or "causes," and therefore their intrinsic natures,[23] are rather different.[24] While the circle in the Liar Paradox originates from a self-referential statement explicitly involving truth value, the circle in SEP originates from a non–self-referential statement (i.e., (A)) involving knowledge.[25] So we have at least a prima facie reason to think that these are two different paradoxes, not two different versions of the same paradox.

Regarding (b), I actually think that the Liar Paradox is much easier to solve than SEP. Eugene Mills offers what I take to be the correct solution, which is very straightforward: Liar statements like (29) are not ultimately paradoxical but rather self-contradictory and therefore false. Not false and then true and then false, etc.—just plain false (see Mills 1998).[26] The same, however, cannot be said of (A). It is not self-contradictory and therefore (just plain) false because, as the Intuition makes clear, it *can* be true. A surprise exam *can* be given on at least *some* days of the testing period. So another solution is required to show how (A) can be true in the face of the Impossibility Argument, which concludes that (A) must be false. SEP's need for a different kind of solution than Mills's solution to the Liar Paradox is yet a further reason to think that SEP may not be reduced to a version of the Liar Paradox.

5.4 Argument 3: The Epistemic Blindspot Argument

The third argument against the No Friday Argument involves an "epistemic blindspot," a concept that several philosophers, most famously Roy Sorensen, have used to elucidate SEP (see Sorensen 1984, 131–35; 1988, chs. 9, 10).[27] To say that Clever has an *epistemic* blindspot to (A) is to say that:

> (30) Clever is justified in believing (A) if and only if (A) is *not* true (or, equivalently, Clever is *not* justified in believing (A) if and only if (A) *is* true).

Together, the Wright & Sudbury Argument and the Renewed No Friday Argument show that Clever has an epistemic blindspot to (A) on Thursday night, that whatever Clever believes on

148

Thursday night helps to make the opposite the case. On the one hand, the Wright & Sudbury Argument shows that if Clever believes on Thursday night that the No Friday Argument is correct and an exam on Friday will not be a surprise, then it will be a surprise. On the other hand, the Renewed No Friday Argument shows that if Clever believes on Thursday night that the No Friday Argument is incorrect and an exam on Friday will be a surprise, then it will not be a surprise.

Of course, Clever is maximally rational and so is arguably *aware* that she has this epistemic blindspot. But the mere fact that she is aware of it does not necessarily dissolve it. Clever is just as susceptible to this blindspot as she would be if she were *not* maximally rational. For while her maximal rationality guarantees that she will not make silly mistakes or invalid inferences, it does *not* guarantee that her (justifiable) belief that a surprise exam will be given on Friday will somehow make the object of this belief true. Again, its being true is *incompatible* with her (justifiably) believing it, and there is nothing she can do to change this situation. Nor can she change the converse—namely, that if she does *not* (justifiably) believe that a surprise exam will be given on Friday, then an exam on Friday *will be* a surprise. In short, Clever's maximally rational powers cannot help her to escape the dilemma between (justifiably) believing a proposition that will then (as a result of her belief) be false and *not* (justifiably) believing a proposition that will then (as a result of her nonbelief) be true. She was inevitably *trapped* in this epistemically unfortunate situation when the teacher announced (A) to her and then did not give her an exam by Thursday, the penultimate day of the testing period.

Clever, then, has no basis, no good reason, for adopting one belief over the other. And given that Clever is maximally rational, she may not arbitrarily choose one belief over the other. Moreover, even if she did have a basis for adopting one belief over the other, the belief would quickly be undermined by the conclusion to which it leads. (Of course, this conclusion would itself be undermined once it was converted into a belief, for it would then lead to the directly opposite conclusion. But this conclusion would also be undermined as soon as it was converted into a belief. And so on.) So it seems that Clever must remain neutral between both beliefs. But to say that Clever must remain neutral between both beliefs is just to say that she is not justified in believing one way or the other (see Wright and Sudbury 1977, 54). And if she is not justified in believing one way or the other, then she is not justified in believing on Thursday night that an exam will take place on Friday.[28] So the No Friday Argument fails once again.

6. Conclusion

The Impossibility Argument consists of two sub-arguments: the No Friday Argument and the Backtracking Argument. The No Friday Argument says that a surprise exam cannot be given on the last day of the testing period. The Backtracking Argument assumes that the No Friday Argument is correct and proceeds to eliminate each earlier day of the testing period as a possible day for the exam as well. So the conclusion of the Impossibility Argument is that, contrary to our intuition that a surprise exam *can* be given on at least some days of the testing period, a surprise exam *cannot* be given on *any* day of the testing period.

But this conclusion is false. Our initial intuition that a surprise exam can be given on most, if not all, days of the testing period is correct.

The main task in solving SEP is to show *why* the No Friday Argument fails. And this task itself splits into two sub-tasks. The first sub-task is to explain away the reasons why the No Friday Argument initially strikes us as so convincing. The second sub-task is actually to *refute* the No Friday Argument.

The Misleading Intuitions Argument is designed to accomplish the first task. It maintains that the No Friday Argument seems very convincing at first because it relies on two powerful intuitions. The first intuition is that an exam on Friday will not be a surprise if the student *expects* it on Thursday night. The second intuition is that an exam on Friday will not be a surprise if the student is *certain* of it on Thursday night. Both intuitions, however, fail to do what they initially seem to—namely, support the No Friday Argument. For according to the strongest version of SEP, an exam on Friday is a surprise if and only if the student did not *justifiably believe* on Thursday night that the exam would take place on Friday. And neither mere expectation nor psychological certainty is sufficient for justified belief.

Again, the Misleading Intuitions Argument does not refute the No Friday Argument; it does not show that the student does *not* justifiably believe on Thursday night that there will be an exam on Friday. Again, it shows only that our likely reasons for initially "giving a pass" to the No Friday Argument are weak. And by exposing the confusion behind our initial assent to the No Friday Argument, it helps to make the three arguments designed to *refute* the No Friday Argument in section 5 that much easier to accept.

The first of these three arguments is the Unprojectible Announcement Argument. The Unprojectible Announcement Argument helps to show that an assumption on which the No Friday Argument tacitly depends—namely, that the teacher's announcement that there will be a surprise exam by the end of the testing period may be applied unproblematically to the last night of the testing period, Thursday night—is false. This assump-

150

tion is false because the student's interpretation of the announcement must change when she applies it to Thursday night. It must shift from "There will be a surprise exam by the end of the testing period" to "There will be a surprise exam *tomorrow*." And this latter statement is difficult to make sense of. The part of this statement that there will be an exam on Friday justifies the student's belief that there will be an exam on Friday. But the part of this statement that the student will be *surprised* suggests the very opposite—that the student is *not* justified in believing that there will be an exam on Friday. This inconsistency suggests that it is irrational for the student to accept the No Friday Argument's tacit assumption about the teacher's announcement.

Second, the Wright & Sudbury Argument constitutes a reductio ad absurdum of the No Friday Argument. The Wright & Sudbury Argument suggests that if the No Friday Argument is correct that an exam on Friday will not be a surprise, then the student is no longer justified in believing on Thursday night the teacher's announcement, in which case an exam on Friday *will* be a surprise.

Third, the No Friday Argument is vulnerable to the Epistemic Blindspot Argument. It may be shown that the Wright & Sudbury Argument together with another argument in support of the conclusion of the No Friday Argument—what I have called the Renewed No Friday Argument—create an epistemic blindspot. On the one hand, the Wright & Sudbury Argument helps to show that if the student believes on Thursday night that the teacher can no longer fulfill his announcement and give a surprise exam, then an exam on Friday *will* be a surprise. On the other hand, the Renewed No Friday Argument helps to show that if the student believes on Thursday night that a surprise exam *will* be given on Friday, then it will *not* be a surprise. Because the student is maximally rational—another assumption that the strongest version of the Surprise Exam Paradox requires us to make —and because she has no good reason to adopt one belief over the other, she must remain neutral between both beliefs. And to say that she must remain neutral between both beliefs is just to say that she may not *justifiably* adopt either. Because one of the two beliefs that the student may not justifiably adopt is that the exam will be given on Friday, the No Friday Argument fails. For, once again, the No Friday Argument concludes that the student *does* justifiably believe on Thursday night that the exam will be given on Friday.

Because the No Friday Argument fails several times over, the Impossibility Argument fails several times over. And because the Impossibility Argument fails, our intuition that a surprise exam *is* indeed possible has been vindicated. A surprise exam *may* be given on any day of the testing period, including the very last day.

Notes

[1] Other philosophers who make the same assumption include Binkley (1968, 127–29), Cargile (1967, 556), and McLelland and Chihara (1975, 74, 76). Cargile (1967, 559–60) argues that the *teacher* must also be maximally rational and that these "idealizing assumptions" are what ultimately generate SEP in the first place. See also Sorensen 1984, 131. Weintraub (1995, 167) and Williamson (1992, 230) reject the assumption that Clever must be maximally rational. Harrison (1975, 74, 80–81) suggests that while this assumption is helpful, it is not necessary. Kirkham (1991, 37) does not object to this assumption "in principle" but does complain that "some writers have idealized the agents in question to the point that their versions fail to describe a *prima facie* counterexample to logic that is of any interest."

[2] Hall (1999, 650–51) spells out these conditions of rationality in more detail.

[3] It follows from this definition of surprise that only the occurrence of an exam may be a surprise, not the absence of an exam. To be sure, if the teacher gave no exam at all during the testing period, Clever would likely be surprised in *some* sense. But, again, the sense in which she would be surprised would not be the sense in which surprise is defined here. See Ferguson 1991, 299.

[4] My formulation of the Impossibility Argument does not explicitly or implicitly depend on the "KK Axiom" or "KK Principle"—i.e., the principle that if Clever knows X, then she knows that she knows X. (Discussion of the KK Principle originated in Hintikka 1962.) If I am right about this point, then Harrison (1975), McLelland (1971), and McLelland and Chihara (1975) fail to offer the correct solution to SEP. For all suggest that the problem with the Impossibility Argument is that it rests on the KK Principle and that the KK Principle is false (at least in this context). Sharvy (1983) uses the "Bottle Imp Paradox" to reach the same conclusion. See also Kirkham 1991, 39; Sorensen 1982, 358–60; and Wright and Sudbury 1977, 46. Williamson (1992, 230) agrees that the Impossibility Argument does not necessarily rest on the KK Principle. But he does think that the Impossibility Argument suffers from the "failure of the KK Principle," which he suggests is both "natural" and "systematic," as well as from Clever's related "inexactness of knowledge" (1992, 224–26, 231). See also Williamson 2000, 140–42. Hall (1999, 651) offers a principle ("Introspection") similar to the KK Principle, but it is framed in terms of justified belief rather than knowledge.

[5] Ferguson (1991), Guiasu (1987), and Meltzer and Good (1965) all hold that the conflict between the Impossibility Argument and the

Intuition is only apparent, not real. For, despite first appearances, the Impossibility Argument involves one sense of surprise, the Intuition another. Incidentally, there is disagreement among these authors (not, of course, between co-authors Meltzer and Good) about what these different senses of surprise actually are.

[6] Lewis (1984, 143) makes a similar point: "You don't make a paradox go away by talking about an unparadoxical case instead."

[7] See Champlin 1976, Chihara 1985, Kaplan and Montague 1960, McLelland 1971, McLelland and Chihara 1975, Quine 1953, Sorensen 1984, and Williamson 1992. Philosophers who formulate SEP in terms of justified belief rather than knowledge/justified-and-true belief include Cargile (1967), Galle (1981), Hall (1999, 648), Nielsen (1979), Olin (1983, 1986), Schoenberg (1966), and Wright and Sudbury (1977, 44, 48–49). Williams (2007) considers both formulations, justified belief and justified-and-true belief. Hall (1999, 695) suggests, correctly I think, that it makes little difference to SEP whether we formulate it in terms of justified belief or justified-and-true belief.

[8] I will treat *justified belief* and *justified in believing* as equivalent. Since Clever is maximally rational, she believes a given proposition if and only if she is justified in believing the proposition. See Hall 1999, 648; Wright and Sudbury 1977, 52–53; and Williams 2007, 91 n. 9.

[9] Quine's solution (1953) depends on the assumption that (A) may *not* be true, that the teacher may *not* fulfill (A) even if he can. But I will argue in section 2.9 that the opposite assumption (i.e., that the teacher will fulfill (A) if he can) is essential to the strongest version of SEP. If I am right about this, then Quine succeeds at best only in solving a weaker version of SEP. Philosophers who also arrive at a negative assessment of Quine's solution include Ayer (1973, 125), Chihara (1985, 196), Edman (1974, 168), Hall (1999, 659ff.), Janaway (1989, 392), Sorensen (1984, 129), and Wright and Sudbury (1977, 42–43).

[10] One might argue against (3) that Clever has *independent* reasons for adopting either conjunct. But she can't, at least not for the purposes of SEP. For without (3), the paradox in SEP would vanish. If Clever had independent reason on any given night to believe that there would be an exam the next day, then it seems fairly straightforward and unproblematic that an exam the next day would *not* surprise her. See Olin 1983, 230–31. So, again, her only reason for adopting either conjunct—$(A)_E$ or $(A)_S$—must be the teacher's announcement of this conjunction—i.e., (A); in which case, if the teacher had not announced (A), Clever would have *no* reason to believe that a surprise exam is coming.

[11] Notice that if (A) and the Impossibility Argument were formulated in terms of belief rather than in terms of *justified* belief, the Impossibility Argument would fare much better. For then the objection that I just raised against it would be unsuccessful. Still, (A) and the Impossibility Argument may not be formulated in terms of belief for the reasons that I gave toward the end of section 2.7.

[12] Jongeling and Koetsier (1993, 308) go so far as to suggest that there is "no basic difference between the [many-day] version and the [one-day] version."

[13] Janaway (1989) thinks that the Impossibility Argument cannot be correct because it rests on "incoherent reasoning." The reasoning is incoherent because it starts out assuming one thing—namely, that Clever knows that she will be surprised by an exam—and ends up

concluding the opposite—namely, that Clever cannot be surprised by an exam. Two problems with Janaway's argument are that he may just be restating the paradox and that the Impossibility Argument as Janaway characterizes it has the form of a reductio ad absurdum, which is a perfectly coherent kind of reasoning.

[14] See Gardner 1963, 154; Goldstein 1993, 96; Jongeling and Koetsier 1993; Kiefer and Ellison 1965; Margalit and Bar-Hillel 1984, 284–85; Nerlich 1961, 513; Olin 1983, 228–29; Slater 1974, 50; Williams 2007, 74; Williamson 2000, 138; and Wright and Sudbury 1977, 50.

[15] Philosophers who draw a connection between SEP and Moore's Paradox include Binkley (1968, 135), Goldstein (1993), and Wright and Sudbury (1977, 46–47, 49).

[16] My worry about classifying this proposition-type, which represents every leap of faith, as Moore-Paradoxical is that this label implies irrational, irrational implies undesirable, and yet some leaps of faith—e.g., some religious beliefs—are arguably desirable.

[17] It is not clear to me that Almeida's later explanation succeeds in showing that the other version of (11)—i.e., p and I do not justifiably believe p (as opposed to p and I justifiably believe $\sim p$)—is strongly inconsistent and therefore irrational to believe.

[18] Proponents of different versions of the Pragmatic Solution include Champlin (1976, 350–51), Levi (2000), Margalit and Bar-Hillel (1984, 286), and Weintraub (1995). Ferguson's solution (1991, esp. 300–01) is comparable to the Pragmatic Solution insofar as it offers an interpretation of *surprise* that plausibly corresponds with what a "normal" teacher would very likely have intended when he announced (A).

[19] In addition to Wright and Sudbury, philosophers who hold that the epistemological status of Clever's belief in (A) changes for the worse as the last day or two of the testing period approach include Binkley (1968, 136), Kirkham (1991, 37–38, 41ff.), Kripke (unpublished, 9ff.), McLelland (1971, 83–84), Nielsen (1979), Rescher (2001, 113–14), Schoenberg (1966), Williams (2007, 76–77, 80), and Williamson (1992, 229). See Hall (1999, 660, 663ff., 684) for a partially sympathetic, partially critical response to Wright and Sudbury. In opposition to Wright and Sudbury 1977 and Kirkham 1991, see Sorensen 1982, 357–60.

[20] The argument for (14) is similar to the Unprojectible Announcement Argument insofar as both arguments conclude that it is no longer tenable for Clever to believe (A) on Thursday night.

[21] Gardner (1963, 154) and Olin (1983, 228–31; 1986, 182) also endorse the inference from (14) to (18).

[22] For a helpful discussion of vicious and virtuous circles, see Cling 2002 and 2003.

[23] See Priest 2000, 125: "[T]he appropriate level at which to analyse a phenomenon is the level which locates underlying causes.... [T]he correct level of abstraction for an analysis of the paradoxes of self-reference is ... the level of the underlying structure that generates and causes the contradictions."

[24] Contrary to Cargile 1965, Kaplan and Montague 1960, Margalit and Bar-Hillel 1984, 286, and Windt 1973.

[25] There are two different categories of philosophers who would disagree with this statement. In the first category are Tennant (1995) and Yablo (1993). Both argue that self-reference is not essential to the Liar Paradox. Priest (1997), however, disagrees. For a reply to Priest, see Sorensen 1998. For a reply to Sorensen, see Beall 2001b. In the

second category are philosophers who believe that the source of SEP is the fact that (A) is a (disguised) self-referential proposition. Proponents of this "self-referential solution" to SEP include Edman (1974), Gardner (1962), Medlin (1964), Sainsbury (1988), and Shaw (1958). Toward the end of their article, Meltzer and Good (1965, 51) also seem to be proposing this kind of view even though they do not seem to recognize it as such. Philosophers who argue, correctly in my view, that self-referential propositions are not essential to SEP include Sorensen (1984, 1988) and Williamson (1992; 2000, 139).

[26] For other solutions to the Liar Paradox, see Beall 2001a; Bhave 1992; Burge 1979; Goldstein 1985, 1986, 2000, 2001; Greenough 2001; Gupta 1982; Kripke 1975; Martin 1967; Priest 1994; Tarski 1944; and Weir 2000. For a discussion of Tarski's approach to the Liar Paradox, see Sainsbury 1995, 105–15. For a non–solution-oriented approach to semantic paradoxes in general, see Herzberger 1982.

[27] See also Binkley 1968, 130, 134–35; Chapman and Butler 1965; Gardner 1963, 154; Gerbrandy 2007, 25–26; Kiefer and Ellison 1965, 426–27; Kvart 1978; McLelland 1971, 84–85; O'Beirne 1961, 464; Olin 1983, 1986; Williams 2007, 72, 74–80; and Woodall 1967. Cargile (1965, 103) sees a parallel between the student's epistemic blindspot and different versions of the Liar Paradox, which I discussed in section 5.3. See also Jongeling and Koetsier 1993, 301–03. Janaway (1989, 405) seems to think that the student has an epistemic blindspot when he says, "What is perhaps surprising about [SEP] is that a simple prediction, which is rationally and sincerely made, can be seen to be possibly true if and only if it is regarded as possibly false." But Janaway (1989, 398–99) still raises several objections against Olin with respect to her adoption of this position.

[28] Clever is also not justified in believing the other way—i.e., that the exam will *not* be on Friday. Importantly, this point—that Clever is not justified in believing that the exam will not be on Friday—is perfectly consistent with Clever's being surprised by an exam on Friday. The only proposition that is *not* consistent with Clever's being surprised by an exam on Friday is that Clever is *justified* in believing that there will be an exam on Friday. And there is no warrant for *this* proposition. It certainly does not follow from Clever's epistemic blindspot.

References

Austin, A. K. 1969. On the unexpected examination. *Mind* 78:137.
———. 1979. The unexpected examination. *Analysis* 39:63–64.
Ayer, A. J. 1973. On a supposed antinomy. *Mind* 82:125–26.
Beall, J. C. 2001a. A neglected deflationist approach to the liar. *Analysis* 61:126–29.
———. 2001b. Is Yablo's paradox non-circular? *Analysis* 61:176–87.
Bennett, J. 1965. Review. *Journal of Symbolic Logic* 30:101–02.
Bhave, S. V. 1992. The liar paradox and many-valued logic. *Philosophical Quarterly* 42:465–79.
Binkley, R. 1968. The surprise exam in modal logic. *Journal of Philosophy* 65:127–36.
Burge, T. 1979. Semantical paradox. *Journal of Philosophy* 76:169–98.
Cargile, J. 1965. Review. *Journal of Symbolic Logic* 30:102–03.
———. 1967. The surprise test paradox. *Journal of Philosophy* 64:550–63.

Cave, P. 2004. Reeling and A-reasoning: Surprise examinations and Newcomb's tale. *Philosophy* 79:609–16.

Champlin, T. S. 1976. Quine's judge. *Philosophical Studies* 29:349–52.

Chapman, J. M., and R. J. Butler. 1965. On Quine's 'so-called' paradox. *Mind* 74:424–25.

Chihara, C. S. 1985. Olin, Quine, and the surprise examination. *Philosophical Studies* 196:191–99.

Cling, A. D. 2002. Justification-affording circular arguments. *Philosophical Studies* 111:251–75.

———. 2003. Self-supporting arguments. *Philosophy & Phenomenological Research* 66:279–303.

Cohen, L. J. 1950. Mr. O'Connor's 'pragmatic paradoxes'. *Mind* 59:85–87.

de Almeida, C. 2001. What Moore's paradox is about. *Philosophy & 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 Green and John N. Williams. Oxford: Oxford University Press.

Edman, M. 1974. The prediction paradox. *Theoria* 40:166–75.

Ferguson, K. G. 1991. Equivocation in the surprise exam paradox. *Southern Journal of Philosophy* 29:291–302.

Fitch, F. B. 1964. A Goedelized conception of the prediction paradox. *American Philosophical Quarterly* 1:161–64.

Galle, P. 1981. Another note on the 'surprise test' puzzle. *Informal Logic* 3:21–22.

Gardner, M. 1962. A new prediction paradox. *British Journal for the Philosophy of Science* 13:51.

———. 1963, March. A new paradox, and variations on it, about a man condemned to be hanged. *Scientific American* 208:144–54.

Gerbrandy, J. 2007. The surprise examination in dynamic epistemic logic. *Synthese* 155:21–33.

Goldstein, L. 1985. The paradox of the liar—A case of mistaken identity. *Analysis* 45:9–13.

———. 1986. False stipulation and semantical paradox. *Analysis* 86:192–95.

———. 1993. Inescapable surprises and acquirable intentions. *Analysis* 53:93–99.

———. 2000. A unified solution to some paradoxes. *Proceedings of the Aristotelian Society* 100:53–74.

———. 2001. Truth-bearers and the liar—A reply to Alan Weir. *Analysis* 61:115–26.

Greenough, P. 2001. Free assumptions and the liar paradox. *American Philosophical Quarterly* 38:115–35.

Guiasu, S. 1987. Prediction paradox revisited. *Logique et Analyse* 30:147–54.

Gupta, A. 1982. Truth and paradox. *Journal of Philosophical Logic* 11:1–60.

Hall, N. 1999. How to set a surprise exam. *Mind* 108:647–703.

Halpern, J. Y., and Y. Moses. 1986. Taken by surprise: The paradox of the surprise test revisited. *Journal of Philosophical Logic* 15:281–304.

Harrison, C. 1975. The unanticipated examination in view of Kripke's semantics for modal logic. In *Philosophical Logic*, ed. J. W. Davis, D. J. Hockney, and W. K. Wilson. Dordrecht: D. Reidel.

Herzberger, H. 1982. Naïve semantics and the liar paradox. *Journal of

Philosophy 79:479–97.

Hintikka, J. 1962. *Knowledge and belief*. Ithaca: Cornell Press.

Janaway, C. 1989. Knowing about surprises: A supposed antimony revisited. *Mind* 98:391–409.

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.

Kiefer, J., and J. Ellison. 1965. The prediction paradox again. *Mind* 74:426–27.

Kirkham, R. 1991. On paradoxes and a surprise exam. *Philosophia* 21:31–51.

Klein, P. D. 1998. Certainty. In *Routledge Encyclopedia of Philosophy*, ed. E. Craig. London: Routledge.

Kripke, S. 1971. Two paradoxes of knowledge (unpublished). Lecture given to the Moral Sciences Club, Cambridge.

———. 1975. Outline of a theory of truth. *Journal of Philosophy* 72:690–715.

Kvart, I. 1978. The paradox of the surprise examination. *Logique et Analyse* 21:337–44.

Levi, D. S. 2000. Surprise! *Southern Journal of Philosophy* 38:447–64.

Lewis, D. 1984. Devil's bargains and the real world. In MacLean 1984, 141–54.

Lyon, A. 1959. The prediction paradox. *Mind* 68:510–17.

MacLean, D., ed. 1984. *The security gamble*. Totowa, NJ: Rowman & Allanheld Publishers.

Margalit, A., and M. Bar-Hillel. 1984. Expecting the unexpected. *Philosophia* 13:263–88.

Martin, R. L. 1967. Toward a solution to the liar paradox. *Philosophical Review* 76:279–311.

McLelland, J. 1971. Epistemic logic and the paradox of the surprise examination. *International Logic Review* 3:69–85.

McLelland, J., and C. Chihara. 1975. The surprise examination paradox. *Journal of Philosophical Logic* 4:71–89.

Medlin, B. 1964. The unexpected examination. *American Philosophical Quarterly* 1:66–72.

Meltzer, B. 1964. The third possibility. *Mind* 73:430–33.

Meltzer, B., and I. J. Good. 1965. 2 forms of the prediction paradox. *British Journal for the Philosophy of Science* 16:50–51.

Mills, E. 1998. A simple solution to the liar. *Philosophical Studies* 89:197–212.

Nerlich, G. C. 1961. Unexpected examinations and unprovable statements. *Mind* 70:503–13.

Nielsen, H. 1979. A note on the 'surprise test' puzzle. *Informal Logic* 2:6–7.

O'Beirne, T. H. 1961. Can the unexpected *never* happen? *New Scientist* 10:464–65.

O'Connor, D. J. 1948. Pragmatic paradoxes. *Mind* 57:358–59.

Olin, D. 1983. The prediction paradox resolved. *Philosophical Studies* 44:225–33.

———. 1986. The prediction paradox: Resolving recalcitrant variations. *Australasian Journal of Philosophy* 64:181–89.

———. 1988. Predictions, intentions and the prisoner's dilemma. *Philosophical Quarterly* 38:111–16.

Priest, G. 1994. The structure of the paradoxes of self-reference. *Mind* 103:25–34.

———. 1997. Yablo's paradox. *Analysis* 57:236–42.

———. 2000. On the principle of uniform solution: A reply to Smith. *Mind* 109:123–26.

Quine, W. V. 1953. On a so-called paradox. *Mind* 62:65–67. Reprinted in *Ways of paradox and other essays*, 21–23. New York: Random House, 1961.

Rescher, N. 2001. *Paradoxes: Their roots, range, and resolution.* Chicago and La Salle, IL: Open Court.

Sainsbury, M. 1988. *Paradoxes.* Cambridge: Cambridge University Press.

———. 1995. Philosophical logic. In *Philosophy: A guide through the subject*, ed. A. C. Grayling, 61–122. New York: Oxford University Press.

Schick, F. 2000. Surprise, self-knowledge, and commonality. *Journal of Philosophy* 97:440–53.

Schoenberg, J. 1966. A note on the logical fallacy in the paradox of the unexpected examination. *Mind* 75:125–27.

Scriven, M. 1951. Paradoxical announcements. *Mind* 60:403–07.

Sharpe, R. A. 1965. The unexpected examination. *Mind* 74:255.

Sharvy, R. 1983. The bottle imp. *Philosophia* 12:401.

Shaw, R. 1958. The paradox of the unexpected examination. *Mind* 67:382–84.

Slater, B. H. 1974. The examiner examined. *Analysis* 35:49–50.

Sorensen, R. A. 1982. Recalcitrant variations of the prediction paradox. *Australasian Journal of Philosophy* 60:355–62.

———. 1984. Conditional blindspots and the knowledge squeeze: A solution to the prediction paradox. *Australasian Journal of Philosophy* 62:126–35.

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

———. 1998. Yablo's paradox and kindred infinite liars. *Mind* 107:137–55.

Tarski, A. 1944. The semantic concept of truth and the foundations of semantics. *Philosophy and Phenomenological Research* 4:341–76.

Tennant, N. 1995. On paradox without self-reference. *Analysis* 55:199–207.

Weintraub, R. 1995. Practical solutions to the surprise-exam paradox. *Ratio* (New Series) 7:161–69.

Weir, A. 2000. Token relativism and the liar. *Analysis* 60:156–70.

Williams, J. N. 2007. The surprise exam paradox: Disentangling two reductios. *Journal of Philosophical Research* 32:67–94.

Williamson, T. 1992. Inexact knowledge. *Mind* 101:217–42.

———. 2000. *Knowledge and its limits.* Oxford: Oxford University Press.

Windt, P. Y. 1973. The liar in the prediction paradox. *American Philosophical Quarterly* 10:65–68.

Woodall, D. R. 1967. The paradox of the surprise examination. *Eureka* 30:31–32.

Wright, C., and A. Sudbury. 1977. The paradox of the unexpected examination. *Australasian Journal of Philosophy* 55:41–58.

Yablo, S. 1993. Paradox without self-reference. *Analysis* 53:251–52.