A User's Guide to the Surprise Exam Paradoxes[1]

John Earman
Dept. of History and Philosophy of Science
University of Pittsburgh

The surprise exam paradox and its variants have achieved zombie-like status in the philosophical literature: despite many attempts to kill them they live on. Some of the most prominent readings of the surprise exam announcement are surveyed. The versions pushed by the logicians are chosen to highlight features of the concept of provability. In this they succeed but at the price of providing contorted self-referential readings of the announcement. The versions pushed by the epistemologists are chosen to provide a stress test for the concepts of knowledge and justified belief. In this they succeed but at the price of entangling the resolution of the paradox with controversies surrounding these concepts. A reading that is free of such controversies and that allows a resolution of the paradox to stand out is offered. This resolution does not provide any deep lessons that could not be learned from other sources. Nevertheless the paradox and its variants deserve to live on as a superb teaching instrument.

1 Introduction

A teacher announces to the students in her Philosophy 101 course that the following week they will be given an exam during class on one of the meeting days, Monday, Tuesday, Wednesday, Thursday, or Friday. And because the teacher wants to keep the students on their toes she also announces that it will be a surprise exam—the students will not know in advance when the exam will take place. Waiting until after class so as not to embarrass

[1] The aficionados of the surprise exam paradox will find little new in this paper. But I hope that my framing of the issues and critical commentary will serve as a useful guide to those who are about to plunge into a large and tangled literature in search of an understanding of why this paradox has resisted a commonly accepted resolution.

her teacher, Laura, a budding logician, politely informs her instructor that, however she tries to arrange it, her plans for a surprise exam cannot succeed. The exam cannot be set on Friday, for it were the students would know on Thursday evening that Friday is the exam day and so will not be surprised. Having ruled out Friday, Thursday can also be ruled out since on Wednesday evening the students will know that Thursday is the exam day. Etc.[2]

There are many names and many variants of the surprise exam paradox (the unexpected hanging, the unexpected tiger, the prediction paradox, ...). One of the most charming and useful ones is the spatialized version introduced by Roy Sorensen (1982, 1984). The teacher asks for five volunteers and, mirable dictu, the volunteers are named Monday, Tuesday, Wednesday, Thursday, and Friday. The teacher lines them up, all facing in one direction with Monday at the front, and Friday at the rear. The line is staggered so that each of the students has a view of the backs of all the students in front of her. The teacher announces that she will put a Post-it on the back of each of the students. One, and only one, of the Post-its has a gold star on it. And, the teacher asserts, the student graced by the gold star will be surprised—she will not know she has received the star until she retrieves the Post-it from her back or the line breaks up and she can see the backs of the other students. Before the teacher can proceed with the demonstration the exceptionally precocious student volunteers inform her that she cannot possibly succeed in doing what her announcement claims. Student Friday, at the back of the line, reasons that the star cannot be placed on his back because if it were he would be able to see that none of the other four have it on their backs and from this infer that it must be on his. Student Thursday can go through the same reasoning, and having ruled out student Friday as wearing the gold star, student Thursday can conclude by similar reasoning that the star cannot be placed on her back. Etc. This spatialized version is useful not only for purposes of demonstration, but it is important in showing that the temporal dimension is not essential to the paradox—the paradox can

[2] The most often repeated origin story attributes the discovery of the paradox to Lennart Ekbom, a Swedish mathematician. The story goes that in 1943 (or 1944?) the Swedish Broadcasting Company announced that there would be a civil defense drill the following week. And to make the drill realistic the day of the drill would not be announced so that the citizens would not know in advance when the drill was to take place. Ekbom detected paradox in the announcement. W. V. O. Quine reports that the paradox was circulating by the end of 1943 among cryptoanalyists in Washington D.C. and that it may have come from Tarski (see Sorensen 1988, pp. 262-263).

arise with information accumulating along a spatial axis rather than along a temporal axis.

Many philosophers have weighed in on the surprise exam paradox and its variants—heavy weights, overweights, and lightweights—with articles appearing in leading philosophy journals as well as the *Notices of the American Mathematical Society* and the *American Mathematical Monthly*.[3] It has found its way into popular culture, as attested by multiple videos on YouTube dramatizing the unexpected hanging version.[4] It is a paradox that refuses to die—someone is always waiting in the wings to claim that the paradox has not been satisfactorily resolved and, perhaps, to claim to offer a new resolution.

There are three mutually reinforcing reasons for the longevity of the surprise exam paradox. One is that the paradox resonates with a number of other paradoxes including the liar, sorities, Moore's paradox, the lottery paradox. Second, the surprise exam is a kind of Rorschach test for philosophy. Logicians see it is an opportunity to display their wares—including Gödel's incompleteness theorems (e.g. Ardeshir and Ramezanian 2012, Chow 1998, Fitch 1964, Halpern and Moses 1968, and Kritchman and Raz 2010).[5] Epistemologists see it as an opportunity to explore the concepts of knowledge and justified belief (e.g. Sorensen 1982, 1984, 1988, and 2017). Still others see it as a hybrid of logical and epistemological issues (e.g. Kaplan and Montague 1960). Third, the variety of reactions to this Rorschach test is fueled by the fact that *the* surprise exam announcement is a misnomer: there are multiple ways of reading the announcement, and the resulting paradoxes, if any, call for resolutions that may differ from reading to reading.

In what follows I will examine some of the most prominent readings of the surprise exam announcement. Broadly speaking, these readings can be divided into two categories: self-referential vs. non-self-referential. Many of

[3] The philosophical discussion apparently commences with O'Conner (1948) followed by a flurry of 1950s articles in *Mind* starting with Scriven (1951). Martin Gardner can take credit for popularizing the paradox; see, for example, Gardner (1963). For a good overview of the tangled history of the philosophical discussion of the surprise exam paradox the reader is referred to Chapter 8 ("History of the Prediction Paradox") of Sorensen (1988).

Chow's (2011) arXix version of his (1998) contains a complete bibliography through 2011; it has nearly 200 entries. And since then the flow of articles continues unabated.

[4] At last count there were 20 or more YouTube videos on the Unexpected Hanging.

[5] The surprise exam paradox has also been analyzed using modal logic (Binkley 1968), dynamic epistemic logic (Gerbrandy 2007), and constructive mathematics (Adeshir and Ramezian 2012).

the self-referential readings produce paradox in the form of antinomies. This makes the paradox resolution simple—the teacher uttered a self-contradictory statement; but at the same time it raises the issue of whether the reading captures the teacher's intention. Other self-referential readings avoid antinomy but only through acts of contortion that produce a surprise exam announcement that most teachers and students wouldn't recognize. The non-self-referential readings are typically framed in terms of knowledge or justified belief. While providing stress-tests for these concepts, the resulting paradoxes and puzzles offer little in the way new insights that are the hallmark of a worthy paradox. Finally, I will propose a deflationary reading of the surprise announcement that the students would take to be natural. It avoids wrangles about how the analyze and how to reason with the concepts of knowledge and justified belief. And it allows the surprise exam paradox to be put quietly to bed.

The striking conclusion that emerges from this critical survey of the literature is that, contrary to popular impression, the problem with the surprise exam paradox is not in finding a resolution but in generating a paradox. Paradox hunters who are determined to find paradox will find one, but only by using implausible readings of the surprise exam announcement or employing suspect assumptions and/or questionable inference principles.

2 The logical turn

Many versions of the paradox take 'surprise' to mean that students will not know on the evening before the exam that the exam will take place the next day. Knowledge is a fraught notion. But whatever its analysis, if the students don't know that X then they cannot prove that X. So the idea behind the logical turn is to avoid controversy about the analysis of knowledge by taking 'surprise' to mean that on the evening before the day of the exam the students will not be able to prove that the exam will fall on the following day. And to lend precision to the enterprise, provability is taken to mean provability in a formal system of inference L. The properties of L will prove to be relevant to the statement and resolution of some forms of the paradox

2.1 First version of the surprise exam announcement

I will adopt the formal nomenclature of Kritchman and Raz (2010) with some modifications in notation. For the N-day version of the announcement the day of the exam is denoted by m where $1 \leq m \leq N$, and the evenings before possible exam days are indexed by j, where $0 \leq j \leq N-1$. $\Pr[\mathcal{I}_j](X)$ means that in the formal system L there is a proof of X from $\mathcal{I}_j$, the information available to the students on the evening of day j.[6] Using this notation a template for the N-day provability version of the surprise exam announcement is given by

$$A : (m \in \{1, 2, ..., N\}) \& \forall_{0 \leq j \leq N-1}[(m = j+1) \to \neg \Pr[\mathcal{I}_j](m = j+1)]$$

I will work mainly with the case we started with, $N = 5$, but other cases as well will be discussed.

To make this template into a definite announcement we need to know what information $\mathcal{I}_j$ is available on the evening of day j and available in what sense. Presumably it is information known by the students to be true on the evening of day j. This is certainly not information known in the provability sense but information known in the sense of ordinary and scientific discourse. So already the logical turn leads to a mismatch in senses of knowledge. Ignore this for the nonce. $\mathcal{I}_j$ may include, for example, the information furnished by memory that the exam has not taken place as of the evening of day j. And it may also include the information that, surprising or not, there will be an exam the following week. Taking these possibilities on board, the informal form of announcement becomes:

> Attention students! Next week there will be an exam on exactly one of the days Monday, Tuesday, Wednesday, Thursday, or Friday. And on the evening before the exam you will not be able prove, on the basis of the fact that the exam will be on exactly one of the days Monday, Tuesday, Wednesday, Thursday, or Friday and on the basis of the information that the exam has not taken place as of that evening, that the exam will be on the next day.

[6] Strictly speaking there should be quotation marks around X to indicate that it a sentence of the language of the system L. But since the notation is complicated enough already I will omit the quotation marks.

Using the suggested notation the formalization of the five-day version of the announcement is:

$$A_1^5 :\ (m \in \{1,2,3,4,5\})\&\forall_{0\leq j\leq 4}[(m=j{+}1) \rightarrow \neg\,\mathrm{Pr}[m \in \{1,2,3,4,5\}, m \notin \{0,..,j\}](m=j{+}1)]$$

Pr can be taken to be the proof predicate for calculus (PC), which is provably consistent and complete. (A_1 is a sentence of some suitable meta-language for PC, say, ZFC where the proof predicate Pr[...](...) for PC is defined. But the point is that the sentences being plugged into the square and round brackets in Pr[...](...) can be replaced by well-formed formulas of PC, although I will not bother to do so here.) With this understanding, no paradox results from A_1. Take the case of $j = 4$. A_1^5 implies $(m = 5) \rightarrow \neg\,\mathrm{Pr}[m \in \{1,2,3,4,5\}, m \notin \{0,1,2,3,4\}(m = 5)]$. Obviously $\mathrm{Pr}[m \in \{1,2,3,4,5\}, m \notin \{0,1,2,3,4\}](m = 5)$ holds. So by modus tollens $m \neq 5$, establishing the base case for the backward induction.[7] But the backwards induction stops there. Take the next case of $j = 3$. A_1^5 implies $(m = 4) \rightarrow \neg\,\mathrm{Pr}[m \in \{1,2,3,4,5\}, m \notin \{0,1,2,3\}](m = 4)]$. But now modus tollens cannot be applied since $\neg\,\mathrm{Pr}[m \in \{1,2,3,4,5\}, m \notin \{0,1,2,3\}](m = 4)](m = 4)$, unless $PC + (m \in \{4,5\})$ is inconsistent—which obviously it is not.

Evidently then A_1^5 is consistent, and A_1^5 is true if the exam is given on any day but Friday. Paradox resolved. Hurrah! But not for long—this was too easy.

2.2 A self-referential version of the surprise announcement

The logicians are determined to find paradox so that they can apply their tool kit. (If you have a hammer ...) Towards this end they investigate a self-referential version of the surprise exam announcement.[8] Informally it says:

[7] Backwards inductive reasoning per se is not suspect. It is widely used in game theory; see Auman (1995). Kritchman and Raz (2010) use Chaitin's incompleteness theorem and backwards induction to give a new proof of Gödel's second incompleteness theorem. It is true, however, that backwards induction is used in a number of paradoxes (see Bovens 1997 and Sobel 1993), and this is what may have given it a bad odor.

[8] Shaw (1958) was apparently the first to discuss self-referential versions.

> Attention students! There will be an exam next week on exactly one of the days Monday, Tuesday, Wednesday, Thursday, or Friday. And on the evening before the exam you will not be able prove, on the basis of this announcement and the information that the exam has not taken place as of that evening, that the exam will be on the next day.

Using the notation introduced above the formalization becomes:

$$A_2^5: \ (m \in \{1,2,3,4,5\}) \& \forall_{0 \leq j \leq 4}[(m = j{+}1) \rightarrow \neg \Pr[A_2^5, m \notin \{0,...,j\}](m = j{+}1)]$$

Self-reference raises some red flags. The most immediate worry is that because of self-reference A_2^5 is not a well formed sentence of the formal system L. But if L is rich enough for arithmetic (e.g. ZFC) then Gödel numbering and diagonalization can be used to overcome this worry (see Chow 1998, Kritchman and Raz 2010, and Shaw 1958 for details). Of course, Gödel used these devices to produce a liar type sentence (saying of itself 'I am not provable in L') belonging to the formal system in order to prove its incompleteness.[9] So one might worry that what we have here is an another form of the liar paradox.[10] In broad a sense that is correct because, like the liar sentence, the self-referential A_2^5 quickly leads to contradiction by backwards induction.

Take the case of $j = 4$. A_2^5 implies $(m = 5) \rightarrow \neg \Pr[A_2^5, m \notin \{0,1,2,3,4\}](m = 5)$. But A_2^5 and $m \notin \{0,1,2,3,4\}$ imply $m = 5$. So $\Pr[A_2^5, m \notin \{0,1,2,3,4\}](m = 5)$ holds, and by modus tollens $m \neq 5$, establishing the base for the backwards induction. Next take the case of $j = 3$. A_2^5 implies $(m = 4) \rightarrow \neg \Pr[A_2^5, m \notin \{0,1,2,3\}](m = 4)$. A_2^5 also implies $m \in \{1,2,3,4,5\}$ and, as we have just seen, A_2^5 implies $m \neq 5$. Thus, $\Pr[A_2^5, m \notin \{0,1,2,3\}](m = 4)$, and by modus tollens $m \neq 4$. Etc. eventuating in the contradiction $(m \in \{1,2,3,4,5\}) \& (m \notin \{1,2,3,4,5\})$.

So *if* A_2^5 is an accurate rendition of the surprise exam announcement then the logicians have succeeded in producing paradox. And since the paradox is in the form of an antinomy where none of the steps in reasoning is invalid, the resolution of the paradox is simply that the teacher has uttered

[9] For a technically correct statement of what the Gödel sentence says see Raatikainen (2020).

[10] For a critical review of attempts to assimilate the surprise exam paradox to the liar, sorites, Moore's paradox, the lottery paradox and others see Sorensen (1988, Chapters 8 and 9).

a contradiction—she seemed to be making a sensible announcement, but taken as an assertion it is equivalent to $X \& \neg X$ for any sentence X. Various commentators have wanted to deny the antecedent: the teacher intended to make an informative announcement so the principle of charity would suggest that a different reading be given to her announcement. Another motivation for wanting a different reading of the announcement is the strange sense of 'know' embodied in A_2^5. The students are deemed to know that the exam is on, say, Friday if they can deduce that the exam is on Friday from A_2^5 and the non-occurrence of the exam on the preceding days. But A_2^5 is inconsistent, leading Halpern and Moses to ask: "[C]an you really be said to know something as a result of deducing it from inconsistent information?" (1986, p. 184). This point prompts a subtle but crucial modification of A_2^5 introduced by Kritchman and Raz (2010).

2.3 Gödel to the rescue (?)

Kritchman and Raz (2010) stick to a self-referential version of the surprise exam announcement, but they propose to modify A_2^5 to reflect the idea that if the students know some evening that the exam is to be held the next day then not only should they be able to prove it will be held the next day but they should not be able to prove that it will be held any other day. With this addendum the formalization of the surprise exam announcement then becomes:

$$
\begin{aligned}
A_3^5 \quad &: \quad (m \in \{1,2,3,4,5\}) \& \forall_{0 \leq j \leq 4}[(m = j+1) \rightarrow \\
\neg\{\Pr[A_3^5, m \quad &\notin \quad \{0,...,j\}](m = j+1) \& \forall_{\substack{1 \leq k \leq 5 \\ k \neq j+1}} \neg \Pr[A_3^5, m \notin \{0,...,j\}](m = k)\}]
\end{aligned}
$$

Antinomy is avoided by this amendment. The $j = 4$ instance of A_3^5 gives $(m = 5) \rightarrow \neg\{\Pr[A_3^5, m \notin \{0,...,4\}](m = 5) \& \forall_{1 \leq k \leq 4} \neg \Pr[A_3^5, m \notin \{0,...,4\}](m = k)\}$. As with A_2^5, A_3^5 also implies $\Pr[A_3^5, m \notin \{0,...,4\}](m = 5)$. But to apply modus tollens to get $m \neq 5$ we also need to satisfy the condition $\forall_{1 \leq k \leq 4} \neg \Pr[A_3^5, m \notin \{0,...,4\}](m = k)$. This condition holds iff the the system $L + A_3^5$ is consistent. So what follows from A_3^5 is not $m \neq 5$ but $Con(L + A_3^5) \rightarrow (m \neq 5)$ (where $Con(L + A_3^5) \equiv \neg \Pr[A_3^5](0 = 1)$). If we could continue in this way to show that for the $j = 3$ case $Con(L + A_3^5) \rightarrow (m \neq 4)$ etc. then we would eventually get $Con(L + A_3^5) \rightarrow (m \notin \{1,2,3,4,5\})$. And since A_3^5 entails $m \in \{1,2,3,4,5\}$ the upshot is $\neg Con(L$

$+ A_3^5$). But as the reader can check, in the $j = 3$ case we cannot show without the help of further assumptions that $Con(L + A_3^5) \rightarrow (m \neq 4)$ and, thus, the backwards induction is halted. What further assumptions would enable the backwards induction? Proving that $Con(L + A_3^5)$ would do the trick; but, as Kritschman and Raz (2010, p. 1458) note, Gödel's second incompleteness theorem shows that $Con(L + A_3^5)$ is not provable in $L + A_3^5$.

One worry here is that whereas the A_2^5 reading of the announcement implies too much (it implies everything) the A_3^5 reading implies too little; for should not a satisfactory reading of the announcement imply that the exam does not occur on Friday?[11] But perhaps A_3^5 is strong enough to account for the students' beliefs about when the exam will take place. For example, the logically adept students may reason as follows: 'I have proved that A_3^5 entails $Con(L + A_3^5) \rightarrow (m \neq 5)$ and, therefore, if I am justified in believing that $Con(L + A_3^5)$ then I am justified in believing $m \neq 5$. And I am justified in believing that $Con(L + A_3^5)$ since I have proved it in a system stronger than $L + A_3^5$. Thus, I am justified in believing $m \neq 5$. But A_3^5 doesn't entail $Con(L + A_3^5) \rightarrow (m \neq i)$ for $i = 1, 2, 3, 4$; or at least I can't prove the entailment, and there seems to be no way to prove it short of proving $\neg Con(L + A_3^5)$, which I have disproved. Hence, while I am justified in believing that the exam will not be on Friday I am free to believe that it will occur any other day.' It is a further question whether or not the students can justifiably believe A_3^5 without reactivating the backwards induction.[12]

This is all very interesting. But the fact that the teacher needs a PhD in mathematical logic (and perhaps also a sadistic streak) to concoct self-referential announcements of the A_3^5 ilk, while the students need at least an MA to understand what the teacher has announced, suggests moving on to consider non-referential readings of the surprise exam announcement that plausibly capture what was intended by the teacher and what was understood by the students in, say, a freshman course on the rationalists.

3 Veering from the logical turn

I have three proposals for how to proceed. First, do not follow the logical turn and replace knowledge with provability. The price to be paid for veering from the logical turn is that the new route drives us into the controversies

[11] This is debatable; see Section 3.1 below.

[12] Questions of this kind will be considered below.

about how to understand the concept of knowledge as used in everyday and scientific discourse. To lower the cost, construe the surprise exam paradox as a paradox about justified belief. Whatever else is required for an agent to know that a proposition is true it certainly includes justified belief; and to say that the students are surprised by the exam day is to imply that the evening before the exam they are not justified in believing that the exam will be on the morrow. Justified belief is also a contested notion, but the contest is easier to manage than for knowledge.[13] Next, for $\mathcal{I}_j$, the information known to be true on the evening of day j, substitute $\mathcal{E}_j$, the evidence available on the evening of day j. This is a vague notion but still clear enough to make progress. For example, $\mathcal{E}_0$, the evidence available on Sunday evening, may consist of the memory that the teacher made a surprise exam announcement on the preceding Friday; memories about the reliability of the teacher's pronouncements; and, perhaps, also the testimony of students who have taken the course previous semesters. And if $m > j$ then $\mathcal{E}_j$ includes the memory that the exam has not occurred as of the evening of day j. These are only examples, and an analysis of what counts as available evidence remains to be given. I have no analysis to offer, so in what follows I will offer a template for the surprise exam announcement. Different ways of filling in the template will have different consequences for paradox. Finally, within this framework I will investigate non-self-referential versions of the surprise exam announcement.

3.1 A justified-belief version of announcement

Informally, the justified belief version of a non-self-referential surprise exam announcement takes the form:

> Attention students! There will be an exam next week on exactly one of the days Monday, Tuesday, Wednesday, Thursday, or Friday. And on the evening before the exam you will not be justified in believing on the basis of the evidence then available to you that the exam will be on the next day.

[13] Eventually I will offer a version of the surprise exam announcement that is not framed in terms of knowledge or justified belief and, thereby, avoids wrangles about how to understand these notions.

Using $JB[\mathcal{E}_j](X)$ to stand for the proposition that on the basis of the evidence $\mathcal{E}_j$ available on the evening of day j the students are justified in believing that X, the formalization of the announcement is

$$A_4^5 : (m \in \{1,2,3,4,5\}) \& \forall_{0 \leq j \leq 4}[(m = j+1) \rightarrow \neg JB[\mathcal{E}_j](m = j+1)]$$

This formulation leaves much unsaid. If the exam has not taken place by day $j > 0$, does the evidence of the students' memories justify them in believing that $m \notin \{1, ..., j\}$? If so, reasoning about A_4^5 should be supplemented by the condition

$$S_1^5 : (m > j > 1) \rightarrow JB[\mathcal{E}_j](m \notin \{1, ..., j\})$$

S_1^5 can, and often is, understood as being part of the announcement, in which case the announcement is given by $\overline{A}_4^5 := A_4^5 \& S_1^5$. Additionally, if $\mathcal{E}_0$ includes evidence that the teacher always carries out her intention to give an exam—whether or not it is a surprise—then we may want to include a second supplementary condition

$$S_2^5 : JB[\mathcal{E}_0](m \in \{1,2,3,4,5\})$$

A reason for taking S_2^5 to be part of the announcement itself is that, intuitively, the one-day version of the announcement is self-contradictory, which it is not if construed as A_4^1 or $\overline{A}_4^1$. For those persuaded of this reasoning the announcement is properly formalized as $\overline{\overline{A}}_4^5 := A_4^5 \& S_1^5 \& S_2^5$.

3.2 Rules of the game

Departing from the logical turn means departing from the clear rules of the road for provability in standard logic. To analyze A_4^5 for its potential to generate paradox we need to know the rules for drawing inferences from statements involving $JB[...](...)$. Here I will consider some principles, variants of which have been employed in the literature on the surprise exam. Before starting it is well to be aware that drawing consequences from principles of reasoning about justified belief that have no backing from a plausible analysis of this concept poses a danger of generating disinformation rather than illumination, a danger that has been widely ignored in the philosophical literature. Nevertheless, it is worth spending some time to see how operating with the these and other principles used in the literature can generate paradox since it may help to explain why it has been thought that the surprise exam announcement harbors paradox.

Here are the principles to be considered:

$$\begin{aligned} JB1 &: \{JB[\mathcal{E}](X)\&JB[\mathcal{E}](Y)\} \rightarrow JB[\mathcal{E}](X\&Y) \\ JB2 &: \{JB[\mathcal{E}](X)\&(\vdash_L X \rightarrow Y)\} \rightarrow JB[\mathcal{E}](Y) \\ JB3 &: \{JB[\mathcal{E}](X)\&(\mathcal{E} \subseteq \widetilde{\mathcal{E}})\} \rightarrow JB[\widetilde{\mathcal{E}}](X) \end{aligned}$$

Here L is a sound system of inference in which the non-self-referential announcement can be formulated. For the A_4^5 formalization of the announcement the predicate calculus suffices for L. But by sacrificing the compact representation achieved in A_4^5, L can be taken to be the sentential calculus, the drawback being that the representing sentential formula may spill off the page for a multi-day version of the announcement. $JB1$ and $JB2$ say respectively that justified belief is closed under conjunction and logical implication. $JB3$ is a no-loss principle for justification as evidence accumulates. More typically in the literature the no-loss principle is stated as no-loss of justified belief (or knowledge) over time. In effect this is covered by $JB3$ if, as will be assumed here, the evidence an agent possesses does not decrease over time. Of course, human agents do suffer memory loss, but such human failings are being ignored here. One virtue of $JB3$ is that it covers non-temporal variants of the paradox, such as Sorensen's spatialed version.

In the literature one also finds in place of $JB2$

$$\begin{aligned} JB? &: \vdash_L X \rightarrow JB[\mathcal{E}](X) \\ JB?? &: \{JB[\mathcal{E}](X)\&JB[\mathcal{E}](X \rightarrow Y)\} \rightarrow JB[\mathcal{E}](Y) \end{aligned}$$

from which $JB2$ can be derived. $JB?$ is certainly acceptable. But, for reasons that will emerge below, $JB??$ is arguably defective, and for this reason I will work with $JB2$.[14]

No sooner are $JB1 - JB3$ written down than questions arise. If X and Y contain occurrences of $JB[-](-)$ then arguably $JB2$ should be strengthened.

[14] $JB1$ and $JB3$ are also arguably defective. The fact that one strand of the literature has been based on principles that are largely or wholly defective is notable. Still, let's proceed in the hope of learning something.

Assume that the base logic L contains modus ponens and let $L_1 := L + JB1, JB2, JB3$ be the logic obtained from L by adding $JB1 - JB3$ as axioms. To say that $JB1 - JB3$ are valid principles of reasoning about justified belief is to say that the system L_1 is sound, which prompts ratcheting $JB2$ up to

$$(JB2)_2 : \{JB[\mathcal{E}](X)\&(\vdash_{L_1} X \to Y)\} \to JB[\mathcal{E}](Y)$$

But again, if $JB1$, $(JB2)_2$, and $JB3$ are valid principles, then it seems that $(JB2)_2$ should be strengthened to

$$(JB2)_3 : \{JB[\mathcal{E}](X)\&(\vdash_{L_2} X \to Y)\} \to JB[\mathcal{E}](Y)$$

where $L_2 := L_1 + (JB2)_2$, etc.. What then is the logic in which entailments of the surprise exam announcement are to be drawn? With $L_n := L_{n-1} + (JB2)_n$, $n = 2, 3, ...$, one could try to define a "limit logic" as $n \to \infty$. This option will not be pursued here. The alternative is to say that there is no fixed surprise exam logic per se; rather, the relevant logic is determined by context: it is L_{n^*} with n^* fixed by however many iterations of $JB2$ are needed to do the desired derivation. Since a proof can contain only a finite number of steps one does not have to worry about defining a "limit logic" because n^* is always finite. In the cases considered below n^* is small. This is messy, but in the absence of a reductive analysis of justified belief that allows the elimination of occurrences of $JB[-](-)$ all one can do is soldier on, hoping that illumination about the surprise exam will emerge.

3.3 Paradox avoided

The first thing to notice is that the surprise exam announcement, read as asserting $\overline{A}_4^5 := A_4^5\&S_1^5$, does not produce an antinomy; indeed, as the reader can check, this version of the announcement does not even entail in any L_n the base case, $m \neq 5$, needed for backwards induction.

However, adding the supplementary condition S_2^5 to give $\overline{\overline{A}}_4^5 := A_4^5\&S_1^5\&S_2^5$ does secure the entailment. To see this consider the case of $j = 4$. The condition S_1^5 reads $(m > 4) \to JB[\mathcal{E}_4](m \notin \{1, ..., 4\})$, and applying $JB3$ to S_2^5 gives $JB[\mathcal{E}_4](m \in \{1, 2, 3, 4, 5\})$.[15] Combining these by $JB1$ and $JB2$ results in $(m > 4) \to (JB[\mathcal{E}_4]((m \notin \{1, ..., 4\})\&(m \in \{1, 2, 3, 4, 5\}))$.

[15] Recall that it is being assumed here that evidence is strictly accumulative over time so that if $l \leq m$ then $\mathcal{E}_l \subseteq \mathcal{E}_m$.

Since $\vdash_L (m \notin \{1, ..., 4\})\&(m \in \{1, 2, 3, 4, 5\}) \rightarrow (m = 5)$ the application of $JB2$ yields $(m > 4) \rightarrow JB[\mathcal{E}_4](m = 5)$. Assume for reductio that $m = 5$. Then by modus ponens $JB[\mathcal{E}_4](m = 5)$. Also from A_4^5 for $j = 4$ we have $(m = 5) \rightarrow \neg JB[\mathcal{E}_4](m = 5)$, and using the reductio assumption and modus ponens gives $\neg JB[\mathcal{E}_4](m = 5)$, contradicting the previously derived $JB[\mathcal{E}_4](m = 5)$. Hence, $\vdash_{L_1} (A_4\&S_1^5\&S_2^5) \rightarrow (m \neq 5)$.

The next thing to notice is that even with the help of the supplementary S_2^5 backwards induction is not engaged. Proceeding as in the case of $j = 4$, the $j = 3$ instances of S_1^5 and S_2^5 yield $(m > 3) \rightarrow JB[\mathcal{E}_3](m \in \{4, 5\})$, and the $j = 3$ instance of A_4^5 gives $(m = 4) \rightarrow \neg JB[\mathcal{E}_3](m = 4)$. So assuming for reductio that $m = 4$ produces $JB[\mathcal{E}_3](m \in \{4, 5\})$ and $\neg JB[\mathcal{E}_3](m = 4)$. To get a contradiction with $\neg JB[\mathcal{E}_3](m = 4)$ so as to conclude by modus tollens that $m \neq 4$ requires also that $JB[\mathcal{E}_3](m \neq 5)$. It as previously established that $\vdash_{L_1} (A_4\&S_1^5\&S_2^5) \rightarrow (m \neq 5)$, but to get from here to $JB[\mathcal{E}_3](m \neq 5)$ requires additional assumptions beyond S_2^5 and/or additional principles of inference.

Let's pause to reflect on what we have learned.

3.4 The YouTube version: 'First we rule out Friday...'

Recall the seductive reasoning of the quick-and-dirty version of the surprise exam paradox: 'First we rule out Friday as the exam day because an exam on that day would not be a surprise. Second, having ruled out Friday we can rule out Thursday on similar grounds. Etc.' If the above analysis is on track then we can see why both moves in the quick-and-dirty version are too quick and too dirty. If the surprise exam announcement is read as $\overline{A}_4^5 := A_4^5\&S_1^5$ the first step falters because the truth of the surprise exam announcement does not "rule out" Friday without the help of an additional assumption about what the students are justified in believing on Thursday evening—e.g. that they are justified in believing $m \in \{1, 2, 3, 4, 5\}$. This seems to be the main point of Quine's (1953) dismissal of the (non-self-referential) surprise exam paradox as a "so-called paradox." But quick dismissal is not warranted if the additional assumption can be a reasonable one, as may be the case here. Indeed, S_2^5 seems so reasonable that that it is sometimes taken to be built into the announcement, and if so then Friday is ruled out, at least if the L_n are the surprise exam logics since (in L_1) $\overline{\overline{A}}_4^5 := A_4^5\&S_1^5\&S_2^5$ implies that

$m \neq 5$.[16] Quine might have added that still further assumptions about what the students are justified in believing and/or still further inference principles are needed in order to move the backwards induction off the base case and towards antinomy. The status of these further assumptions and rules is our next topic.

Before moving on it is well to emphasize that the above discussion was conducted under the presupposition that $JB1 - JB3$ are valid principles for reasoning about justified belief. In fact $JB1$ and $JB3$ are suspect for reasons to be elaborated below. But faulting one or both of them was not needed to dissolve the surprise exam paradox on the current reading of the announcement. What is striking is that, contrary to first appearances, it is not easy to generate paradox from the non-referential surprise exam announcement even when arguably defective principles of reasoning are used.[17] Why then is the reasoning in the quick-and-dirty version of the paradox so seductive? An obvious culprit is an over-interpretation of the sense in which Friday is "ruled out" as the exam day, leading to a misapplication of backwards induction. The truth of the surprise exam announcement, read as $\overline{\overline{A}}_4^5$, does "rule out" Friday in the sense that if $\overline{\overline{A}}_4^5$ is true then the exam cannot be held on Friday. But this does not authorize the students to assume $m \neq 5$ on Wednesday evening so as to "rule out" Thursday. Such an authorization is provided by the self-referential version A_2^5, eventuating in a genuine antinomy. Perhaps those seduced by the the quick-and-dirty version of the paradox are oscillating between the non-self-referential reading $\overline{\overline{A}}_4^5$ and the self-referential A_2^5.

4 Paradox and puzzle redux

Assume that the suprise exam announcement, read as $\overline{\overline{A}}_4^5 := A_4^5 \& S_1^5 \& S_2^5$, is true. Can the students justifiably believe on Sunday that it is true? Some additions to the principles $JB1 - JB3$ that are supposed to govern reasoning about justified belief entail a negative answer. Let's work our way towards identfying those additional principles. Assume for reductio that $A_4^5 \& S_1^5 \& S_2^5$ and $JB[\mathcal{E}_0](A_4^5 \& S_1^5 \& S_2^5)$ are both true. In the preceding section we saw that

[16] A version of this point was part of Ayer's (1973) critical response to Quine (1953).

[17] We will see shortly that faulting the defective principles is needed to resolve more elaborate versions of the paradox.

$\vdash_{L_1} A_4^5 \& S_1^5 \& S_2^5 \rightarrow (m \neq 5)$. Using the first reductio assumption and modus ponens yield the base case, $m \neq 5$, for backwards induction. Furthermore, the assumption of $JB[\mathcal{E}_0](A_4^5 \& S_1^5 \& S_2^5)$ allows the backwards induction to engage. In L_2 combining $\vdash_{L_1} A_4^5 \& S_1^5 \& S_2^5 \rightarrow (m \neq 5)$ and the second reductio assumption, $JB[\mathcal{E}_0](A_4^5 \& S_1^5 \& S_2^5)$, yields $JB[\mathcal{E}_0](m \neq 5)$, and from thence using $JB3$ we arrive at $JB[\mathcal{E}_3](m \neq 5)$. To repeat what was found above, the $j = 3$ instances of S_1^5 and S_2^5 yield $(m > 3) \rightarrow JB[\mathcal{E}_3](m \in \{4, 5\})$, and combining this with $JB[\mathcal{E}_3](m \neq 5)$ by $JB1$ and $JB2$ produces $(m > 3) \rightarrow JB[\mathcal{E}_3](m = 4)$. The $j = 3$ instance of A_4^5 gives $(m = 4) \rightarrow \neg JB[\mathcal{E}_3](m = 4)$. Hence, if $m = 4$ then $JB[\mathcal{E}_3](m = 4) \& \neg JB[\mathcal{E}_3](m = 4)$, and so $m \neq 4$.

So far so good. But the backwards induction falters in the next round. It has been established in the first two rounds that $\vdash_{L_2} ((A_4^5 \& S_1^5 \& S_2^5) \& JB[\mathcal{E}_0](A_4^5 \& S_1^5 \& S_2^5)) \rightarrow (m \notin \{4, 5\})$. If we had the additional assumption that $JB[\mathcal{E}_0](JB[\mathcal{E}_0](A_4^5 \& S_1^5 \& S_2^5))$ then this could be combined by $JB1$ with $JB[\mathcal{E}_0](A_4^5 \& S_1^5 \& S_2^5)$ to get $JB[\mathcal{E}_0]((A_4^5 \& S_1^5 \& S_2^5) \& JB[\mathcal{E}_0](A_4^5 \& S_1^5 \& S_2^5))$, and then in L_3 an application of $(JB2)_3$ yields $JB[\mathcal{E}_0](m \notin \{4, 5\})$. From here it is smooth sailing to the conclusion that $m \neq 3$. The completion of the final two rounds of the backwards induction requires even more assumptions involving more iterated applications of the $JB[-](-)$ operator. Alternatively, to avoid having to add an additional assumption at each new step—which would certainly become tedious in a generalized multi-day version of the paradox—it would suffice to adopt one new inference principle

$$JB4 : JB[\mathcal{E}](X) \rightarrow JB[\mathcal{E}](JB[\mathcal{E}](X))$$

and modify the L_n appropriately to incorporate $JB4$. This is the course I will follow in the remainder of this section.

To summarize, what has been shown so far is that, assuming that inferences about justified belief are carried out in some ($JB4$ modified) logic L_n, if $A_4^5 \& S_1^5 \& S_2^5$ is true then the students cannot be justified in believing on Sunday that it is true, for $A_4^5 \& S_1^5 \& S_2^5$ and $JB[\mathcal{E}_0](A_4^5 \& S_1^5 \& S_2^5)$ together entail a contradiction in L_n. We can now establish a stronger result: If the surprise exam announcement is given the weaker reading $\overline{A}_4^5 := A_4^5 \& S_1^5$ and if $\overline{A}_4^5$ is true then the students cannot be justified in believing on Sunday that $\overline{A}_4^5$ is true. Suppose for reductio that both $A_4^5 \& S_1^5$ and $JB[\mathcal{E}_0](A_4^5 \& S_1^5)$. Since $\vdash_L A_4^5 \& S_1^5 \rightarrow m \in \{1, 2, 3, 4, 5\}$ the assumption $JB[\mathcal{E}_0](A_4^5 \& S_1^5)$ and an application of $JB2$ give $JB[\mathcal{E}_0](m \in \{1, 2, 3, 4, 5\})$, which is just S_2^5; and from thence by $JB4$ we get $JB[\mathcal{E}_0](S_2^5)$. Then combining $JB[\mathcal{E}_0](S_2^5)$ and

$JB[\mathcal{E}_0](A_4^5\&S_1^5)$ by $JB1$ results in $JB[\mathcal{E}_0](A_4^5\&S_1^5\&S_2^5)$. The penultimate conclusion is that the combination of $A_4^5\&S_1^5$ and $JB[\mathcal{E}_0](A_4^5\&S_1^5)$ entails both that $A_4^5\&S_1^5\&S_2^5$ and that $JB[\mathcal{E}_0](A_4^5\&S_1^5\&S_2^5)$. Now we can employ the result of the previous paragraph to get the upshot that the combination of $A_4^5\&S_1^5$ and $JB[\mathcal{E}_0](A_4^5\&S_1^5)$ entails a contradiction in L_n.

The knowledge version of this result is simpler. Let $K[\mathcal{E}](X)$ stand for X is known on the basis of $\mathcal{E}$. Substituting $K[\mathcal{E}](X)$ for $JB[\mathcal{E}](X)$ in A_4^5 gives the knowledge version of the announcement:

$$_KA_4^5 : (m \in \{1,2,3,4,5\})\&\forall_{0\leq j\leq 4}[(m = j+1) \rightarrow \neg K[\mathcal{E}_j](m = j+1)]$$

And under the substitution the supplementary conditions become

$$\begin{aligned} {}_KS_1^5 &: (m > j > 1) \rightarrow K[\mathcal{E}_j](m \notin \{1,...,j\}) \\ {}_KS_2^5 &: K[\mathcal{E}_0](m \in \{1,2,3,4,5\}) \end{aligned}$$

Suppose that inferences about knowledge are governed by the knowledge analogs $K1-K4$ of $JB1-JB4$. Knowledge also obeys another principle

$$K5 : K[\mathcal{E}](X) \rightarrow X$$

since $K5$ is nothing but the codification of the truism that knowledge implies truth. Suppose that the appropriate logics for carrying out inferences inference about knowledge are $_KL_n$, which are defined analogously to the L_n with the principles $JB1-JB4$ replaced by their knowledge analogs $K1-K4$ plus $K5$. By $K5$, if $K[\mathcal{E}_0](_KA_4^5\&_KS_1^5)$ is assumed then $_KA_4^5\&_KS_1^5$ doesn't have to be separately assumed; and transcribing the the above results for justified belief shows that $K[\mathcal{E}_0](_KA_4^5\&_KS_1^5)$ produces antinomy in a $_KL_n$. Results of this type using somewhat different principles governing $K[\mathcal{E}](X)$ can be found in the literature (e.g. Harrison 1969, McLelland and Chihara 1975, Chihara 1985, and Sorensen 1988, pp. 289-290).

These results deserve to be called paradoxes or at least puzzles since intuition suggests that nothing prevents the students on Sunday either from knowing that the surprise exam announcement is true or from being justified in believing the announcement is true if it is indeed true. The resolution of this puzzle favored by many commentators is to reject the KK principle

$$K4 : K[\mathcal{E}](X) \rightarrow K[\mathcal{E}](K[\mathcal{E}](X))$$

and, by implication, the $JBJB$ principle $JB4$ (see for example Harrison 1969 and McLelland and Chihara 1975, 1985). However, this resolution is not fully satisfying. Prima facie, self-reflective epistemologists in possession of correct theories of knowledge and justified belief and self-consciously applying these theories to interesting cases they encounter (such as the surprise exam) will satisfy $K4$ and $JB4$. It is puzzling that these enlightened epistemologists are unable to know that or justifiably believe that the surprise exam announcement is true while their less enlightened brethren can. Perhaps there is something about the surprise exam case that allows enlightened epistemologists to know or be justified in believing the announcement but not to know that they know and not to be justified in believing that they are justified in believing. But that would need to be explained.

I will explore the option of resolving the paradox by rejecting one of the other principles entertained above. But first a few words about blindspots.

4.1 Blindspots

A blindspot for an agent is a contingent proposition that is epistemically inaccessible to that agent. More precisely, a contingent proposition X is a blind spot for an agent whose available evidence is $\mathcal{E}$ just in case, on the basis of $\mathcal{E}$, the agent cannot know that X (knowledge blindspot) or cannot be justified in believing X (justification blindspot), where "cannot" means on pain of contradiction.[18] To illustrate, consider the one-day knowledge version ${}_KA_4^1$ of the surprise exam announcement. It boils down to

$${}_KA_4^1 : (m = 1)\&\neg K[\mathcal{E}_0](m = 1)$$

which is contingent (as is ${}_K\overline{A}_4^1 := {}_KA_4^1\&{}_KS_1^1$ since ${}_K\overline{A}_4^1$ reduces to ${}_KA_4^1$ because ${}_KS_1^1$ is vacuous.[19]) But $K[\mathcal{E}_0](A_4^1)$ is impossible, at least if knowledge obeys $K5$ and distributes across conjunction, i.e.

$$K6 : K[\mathcal{E}](X\&Y) \rightarrow (K[\mathcal{E}](X)\&K[\mathcal{E}](Y))$$

($K6$ is redundant in the presence of $K2$.) Assume for reductio that $K[\mathcal{E}_0]((m = 1)\&\neg K[\mathcal{E}_0](m = 1))$. By $K6$ we get $K[\mathcal{E}_0](m = 1)$ and $K[\mathcal{E}_0](\neg K[\mathcal{E}_0](m =$

[18] The notion of epistemic blindspots was introduced by Roy Sorensen (1982, 1984). Differing definitions of blindspots are offered in Sorensen (1988).

[19] However, ${}_K\overline{\overline{A}}_4^1 = [(m = 1)\&\neg K[\mathcal{E}_0](m = 1)\&K[\mathcal{E}_0](m = 1)]$ is flatly self-contradictory.

1)), and from the latter we get $\neg K[\mathcal{E}_0](m = 1)$ by $K5$. Thus, ${}_K A_4^1$ is a knowledge blindspot for the students on Sunday evening. More interestingly, ${}_K A_4^5 \& {}_K S_1^5 \& {}_K S_2^5$ is a knowledge blindspot for the students on Sunday evening in the proposed knowledge logics.

By contrast the one-day version $A_4^1 : (m = 1) \& \neg JB[\mathcal{E}_0](m = 1)$ of the justified belief form of the announcement (as well as $\overline{A}_4^1 := A_4^1 \& S_1^1$ since $\overline{A}_4^1 = A_4^1$) is not a justification blindspot—$JB[\mathcal{E}_0]((m = 1) \& \neg JB[\mathcal{E}_0](m = 1))$ is consistent since justified belief does not entail truth.[20] However, $\overline{\overline{A}}_4^5 := A_4^5 \& S_1^5 \& S_2^5$ is a *conditional* justification blindspot; for, as we saw above, if $\overline{\overline{A}}_4^5$ is true then $JB[\mathcal{E}_0](\overline{\overline{A}}_4^5)$ cannot be true in the proposed justified belief logics.

Seeing how blind spots can form helps to remove some of the air of paradox surrounding them. But erasing the purported conditional and unconditional epistemic blindspots for the surprise exam can only be accomplished by identifying a false premise or a faulty principle of reasoning, the removal of which restores epistemic accessibility. In the following section I focus suspicion on $JB1$ and $JB3$ and their knowledge counterparts.[21]

5 Puzzle dissolved, blindspots erased (?)

Of the principles $JB1$, $(JB2)_n$, and $JB3$ the least suspect are the $(JB2)_n$. They can fail for students who are not good at logic or simply fail to put two-and-two together, but to cite such failures is an uninteresting way of escaping paradox, especially if paradox is the reward for becoming logic savvy. As will

[20] A first person utterance of $A_4^1 : (m = 1) \& \neg JB[\mathcal{E}_0](m = 1)$ has the form of 'X but I am not justified in believing X'. It has an odd ring not unlike the odd ring of the sentences discussed in Moore's paradox, e.g. 'X but I don't believe X'. For a discussion of attempts to assimilate the surprise exam paradox see Sorensen (1988). A first person utterance of $JB[\mathcal{E}_0]((m = 1) \& \neg JB[\mathcal{E}_0](m = 1))$ sounds even odder. Not only does it sound odd but applying $JB1$, $JB2$, and $JB4$ to it leads to absurdity: by $JB1$, $JB2$, and $JB4$, $JB[\mathcal{E}_0]((m = 1) \& \neg JB[\mathcal{E}_0](m = 1))$ entails $JB[\mathcal{E}_0](JB[\mathcal{E}_0](m = 1) \& \neg JB[\mathcal{E}_0](m = 1))$ and, thus, by $JB2$ it entails $JB[\mathcal{E}_0](Y)$ for any Y whatsoever.

[21] I of course agree with Sorensen that blind spots are implicated in the surprise exam paradox and other epistemic paradoxes. But I do not agree that blindspots lead to a resolution of the paradoxes—they are a statement of the problem rather than part of the resolution; and they get erased by a resolution that consists of identifying a false premise or an invalid step of reasoning that produce the paradox; but see Sorensen (1988, pp. 328-343).

be discussed below, on some analyses of justified belief $JB1$ can fail on more substantive grounds; but these grounds do not seem to apply in the present case. For instance, the use of $JB1$ to combine $JB[\mathcal{E}_4](m \in \{1,2,3,4,5\})$ and $JB[\mathcal{E}_4](m \notin \{1,2,3,4\})$ to give $JB[\mathcal{E}_4]((m \in \{1,2,3,4,5\})\&(m \notin \{1,2,3,4\}))$ and, thereby, $JB[\mathcal{E}_4](m = 5)$ seems defensible. As a general principle $JB3$ is so glaringly defective that it is embarrassing to write it down: an agent may be justified in believing X on the basis of the evidence $\mathcal{E}$ initially available to her, but as additional evidence becomes available the initial evidence in favor of X can be undermined or counterbalanced so that she is no longer justified in believing X on the basis of the total accumulated evidence $\widetilde{\mathcal{E}} \supset \mathcal{E}$ available to her. Various commentators have used a rejection of $JB3$ and the concomitant justification loss that occurs as the days pass without an exam as a means of avoiding the present form of paradox (see, for example, Wright and Sudbury 1977). What needs to be shown is not just that $JB3$ fails as a general principle—which can be taken as a given—but that as applied to the surprise exam its failure explains why neither $A_4^5\&S_1^5\&S_2^5$ by itself nor in conjunction with $JB[\mathcal{E}_0](A_4^5\&S_1^5\&S_2^5)$ produces a contradiction and, thus, why there are no resulting justification blindspots.

In establishing the base case, $m \neq 5$, for the backwards induction the main role for $JB3$ is going from $JB[\mathcal{E}_0](m \in \{1,2,3,4,5\})$ to $JB[\mathcal{E}_4](m \in \{1,2,3,4,5\})$, which is defensible; indeed, unless between Sunday and Thursday evening the students have acquired evidence that the teacher has abandoned her intention to give an exam, it seems that if they are justified in believing on Sunday that there will be an exam, surprise or not, then they are still justified in believing it on Thursday evening.

The second role for $JB3$ lies in engaging the backwards induction and showing that $A_4^5\&S_1^5\&S_2^5$ and $JB[\mathcal{E}_0](A_4^5\&S_1^5\&S_2^5)$ together entail $m \neq 4$. Having established that $A_4^5\&S_1^5\&S_2^5$ entails $m \neq 5$, the assumption $JB[\mathcal{E}_0](A_4^5\&S_1^5\&S_2^5)$ can used to get $JB[\mathcal{E}_0](m \neq 5)$ without employing $JB3$. But $JB3$ is needed in going from $JB[\mathcal{E}_0](m \neq 5)$ to $JB[\mathcal{E}_3](m \neq 5)$. What is illegitimate with this use of $JB3$? Well, one might say, the reason the students initially believed that the exam will not be on Friday derived from their belief in $A_4^5\&S_1^5\&S_2^5$; but after no exam on Monday, Tuesday, or Wednesday that initial belief is undermined so that on Wednesday evening they are no longer justified in believing $A_4^5\&S_1^5\&S_2^5$. If this critique of $JB3$ succeeds in undermining the engagement of the backwards induction it would follow that the students cannot justifiably believe the two-day version $A_4^2\&S_1^2\&S_2^2$ where

$$A_4^2 : (m \in \{1, 2\}) \& \forall_{0 \leq j \leq 1}[(m = j + 1) \rightarrow \neg JB[\mathcal{E}_j](m = j + 1)]$$

and S_1^2 and S_2^2 are the two-day versions of S_1 and S_2:

$$S_1^2 : (m = 2) \rightarrow JB[\mathcal{E}_1](m \neq 1)$$

$$S_2^2 : JB[\mathcal{E}_0](m \in \{1, 2\})$$

Without begging the question of the validity of the principles $JB1$, $(JB2)_n$, and $JB3$ this claim can only be adjudicated if one has at hand an account of when an agent is warranted in claiming justified belief on the basis of the evidence available to her.

With an analysis in hand of justified belief, supplying necessary and sufficient conditions for an agent to be justified in believing X on the basis of evidence $\mathcal{E}$, the principles governing reasoning with justified belief are precisely those that flow from the analysis, no more and no less. Of course, a tug of war commences if these rules conflict with the rules one initially thinks ought to govern reasoning about justified belief, and deciding the winner can be a matter of some nicety. There is another reason for wanting an analysis of justified belief. A fully satisfying resolution of the surprise exam paradoxes and puzzles is not achieved by showing how formal contradiction is avoided. One also would like to know *how* the teacher can set a surprise exam and *how* the students can be justified in initially believing that they will be surprised, even though that justification may fade as the week progresses with no exam having taken place. In the following section I will offer a trial account of justified belief and test to see whether it meets this challenge.

6 How to set a surprise exam, or rather how to set a probably surprising exam

6.1 Justified belief

In the present section I will explore the idea that an agent is justified in believing X on the basis of evidence $\mathcal{E}$ just in case $P(X/\mathcal{E}) \geq p^* > \frac{1}{2}$, where $P(X/\mathcal{E})$ is the agent's rational degree of belief in X conditional on $\mathcal{E}$, and

p^* is some chosen cut-off value. In a later section I will discuss some of the shortcomings of this account of justified belief; but whatever its shortcomings it has the virtue of illustrating the kinds of considerations that must be dealt with if the surprise exam paradox is to be put to rest.

On this analysis our five-day justified belief version of the surprise exam announcement becomes

$$\widehat{A}_4^5 : (m \in \{1,2,3,4,5\})\&\forall_{0\leq j\leq 4}[(m = j+1) \to \neg(P(m = j+1)/\mathcal{E}_j) \geq p^*)]$$

and the supplementary principles become

$$\widehat{S}_1^5 : (m > j > 0) \to P(m \notin \{1,...,j\})/\mathcal{E}_j) \geq p^*$$

$$\widehat{S}_2^5 : P((m \in \{1,2,3,4,5\}))/\mathcal{E}_0) \geq p^*$$

6.2 The inference principles

The inference principles for justified belief considered in Section 3.1 above do not fare well on the cut-off analysis. The conjunction principle $JB1$ fails since whatever the value of the cut-off p^*, as long as it is strictly less than 1, $P(X/\mathcal{E}) \geq p^*$ and $P(Y/\mathcal{E}) \geq p^*$ do not imply $P(X\&Y/\mathcal{E}) \geq p^*$. And the no-loss principle $JB3$ fails miserably since since $P(X/\mathcal{E}) \geq p^*$ does not imply $P(X/\widetilde{\mathcal{E}}) \geq p^*$ for $\mathcal{E} \subseteq \widetilde{\mathcal{E}}$. The problem with the $JBJB$ principle $JB4$ is not that it fails on the cut-off analysis of justified belief but that the cut-off analysis is stymied in providing an evaluation. Such an evaluation would involve evaluating expressions of the form $P(P(X/\mathcal{E}) \geq p^*/\mathcal{E}) = ?$, but the domains of ordinary credence functions do not contain sentences expressing propositions about credences. This might seem to bode ill for assessing whether the students can justifiably believe the surprise exam announcement. I will return to this issue shortly.

Not only does the cut–off analysis serve as a wrecking ball but it also has the more positive virtue of lending itself to an explanation of how the teacher can set a surprise exam.

6.3 Setting the exam[22]

Here I model the students as Bayesian agents who approach the surprise exam paradox with open minded priors; specifically, they assign non-zero, but perhaps widely divergent, priors to the relevant propositions about the to-be-announced exam. But because of two further assumptions their posterior degrees of belief quickly coalesce: first, when the teacher makes the surprise exam announcement $\widehat{A}_4^5$ she also tells the students that the day of the exam will be chosen by a quantum randomizer, and the students take this pronouncement on board; second, the students conform to David Lewis' Principle Principle, to wit, if $\mathcal{E}(x)$ is the evidence that the objective chance of X is x then $P(X/\mathcal{E}(x)) = x$ (see Lewis 1980). Then on Sunday evening the student's rational degree of belief $P(m = i/\mathcal{E}_0)$, conditional on the evidence $\mathcal{E}_0$ then available to them, that the exam will be on day $1 \leq i \leq 5$ is $\frac{1}{5}$ for each i, per Lewis' Principle Principal. Let us also assume that the students' credences obey the supplementary principles $\widehat{S}_1^5$ and $\widehat{S}_2^5$. And to simplify matters, let us guarantee the satisfaction of these conditions by setting $P((m \in \{1, 2, 3, 4, 5\}))/\mathcal{E}_0) = 1$ and $P(m \notin \{1, ..., j\})/\mathcal{E}_j) = 1$ for $j > 0$. Then $P((m = 1))/\mathcal{E}_1) = 0$ if the exam does not occur on Monday. And if the exam does not occur on Monday I will assume that, barring additional evidence that favors one of the remaining days, the rational redistribution of the 1/5 degree of belief initially assigned to Monday is to split it equally among the remaining days so that $P(m = i/\mathcal{E}_1) = \frac{1}{4}$ for $2 \leq i \leq 5$. Proceeding in this way, the rational conditional degrees of belief on successive days are tabulated as

[22] In this section I am relying on Ned Hall's (1999) proposal for how to set a surprise exam. For another proposal on how to set a surprise exam see Sober (1998).

$$\begin{array}{cccccc}
 & M & T & W & Th & F \\
S & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{1}{5} \\
M & 0 & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} & \frac{1}{4} \\
T & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\
W & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\
Th & 0 & 0 & 0 & 0 & 1
\end{array}$$

where the entry in the ith column and jth row is $P(m = i/\mathcal{E}_j)$, $1 \leq i \leq 5$, $0 \leq j \leq 4$, and $\mathcal{E}_j$ is the evidence available to the students on the evening of day j.

Call day i a *surprise day* if, on the basis of the evidence available on the evening of day $i-1$, the students are not justified in believing that $m = i$. On the present interpretation of justified belief this means that $P(m = i/\mathcal{E}_{i-1}) < p^*$. As an illustration set $p^* = 0.8$. Then the surprise days are Monday, Tuesday, Wednesday, and Thursday. The teacher's method for setting the exam day is not infallible in producing a surprise exam—one-fifth of the time it will misfire because the exam will fall on the non-surprise day Friday.[23] What can be said in favor of the proposed randomizing method of setting the exam this: If it is used then four-fifths of the time the exam occurs on Monday, Tuesday, Wednesday, or Thursday, in which case the students will be surprised, and $\widehat{A}_4^5 \& \widehat{S}_1^5 \& \widehat{S}_2^5$ is true.

[23] Can there be an infallible method for setting a surprise exam in the sense that in every possible application of the method the students are surprised? Prima facie it seems not for any form of the exam announcement that implies the exam will not be on Friday. For if such an infallible method were applied then necessarily never on F an exam, so the five-day exam window is reduced to four, making the the five-day announcement equivalent to a four-day announcement. The same consideration applies to the four-day announcement, making it equivalent to a three-day announcement. Etc. However, one knows from the history of the surprise exam announcement that superficially attractive arguments turn out to be fallacious. This one may turn out to fall into that category.

6.4 Blind spots again

Enquiring about justification blindspots on the current analysis of justified belief does not produce an immediate response; for on the cut-off analysis evaluating expressions like $JB[\mathcal{E}_0](\widehat{A}_4^5 \& \widehat{S}_1^5)$ and $JB[\mathcal{E}_0](\widehat{A}_4^5 \& \widehat{S}_1^5 \& \widehat{S}_2^5)$ requires evaluating expressions of the form $P(\widehat{A}_4^5 \& \widehat{S}_1^5 / \mathcal{E}_0) =?$ and $P(\widehat{A}_4^5 \& \widehat{S}_1^5 \& \widehat{S}_2^5 / \mathcal{E}_0) = ?$. But for reasons explained above the domains of ordinary credence functions do not contains expressions of the form $\widehat{A}_4^5 \& \widehat{S}_1^5$ and $\widehat{A}_4^5 \& \widehat{S}_1^5 \& \widehat{S}_2^5$. This seeming inability to deliver a verdict on matters we want to settle is not fatal to the cut–off analysis. What one wants to know is whether the students can justifiably believe on Sunday both that there will be an exam on one of the appointed days and that the day of the exam will be a surprise. In the scheme described above this is functionally equivalent to asking whether the students can justifiably believe on Sunday both that there will be an exam on Monday, Tuesday, Wednesday, Thursday, or Friday and that the exam will fall on one of the surprise days Monday, Tuesday, Wednesday, or Thursday. The scheme above demonstrates that this question can be answered in the affirmative; for according to this scheme, on Sunday their rational credence in the conjunction of these two propositions reaches the chosen cut-off value of 0.8 for justified belief.

6.5 Raising the bar

In the above I rigged the choice of p^* to get the result I wanted. This rigging was justified in order to give a proof-of-concept, but it brings into question how robust the scheme is. Setting the bar higher for justified belief changes the conclusions of the analysis. Suppose, for example, the cut-off is set to $p^* = 0.9$ rather than to 0.8. The surprise days are still Monday, Tuesday, Wednesday, and Thursday. But on Sunday the rational conditional degree of belief that the exam will occur on a surprise day is only 0.8 which falls below the cut-off value, so on Sunday the students are not justified in believing that the exam will fall on a surprise day. This can be overcome by increasing the number N of possible exam days and using a quantum randomizer to give equal weight to each of the N days. The number of surprise days is now $N - 1$, and on Sunday the rational degree of belief that the exam will occur on a surprise day is $\dfrac{N-1}{N}$, which can be made to as close to unity as desired by making N sufficiently large. In this way no matter how high the bar for justified belief is set, as long as p^* is strictly less than 1, the bar can be cleared by making the number of possible exam days sufficiently large.

Raising the bar for justified belief simultaneously lowers the bar for surprise if the exam is defined to be a surprise when on the evening of the day before the exam the students are not justified in believing that the exam will be the next day, and on the present analysis this means that the rational conditional degree of belief that the exam will take place the next day is less than p^*. If p^* is close to unity the analysis does not yield a very appealing notion of surprise. Thus, one might want to redo the analysis with a higher bar for surprise. In the present setting this amounts to counting an exam on day i as surprising iff the rational conditional degree of belief on the evening of day $i-1$ that the exam will take place on i is less than or equal to some $0 \leq s^* < \frac{1}{2}$. I will return to this suggestion in Section 8.

Finally it is worth seeing how the present analysis of justified belief treats the two-day version $\widehat{A}_4^2$ of the surprise exam announcement. Whatever objective chances p_1 and p_2, $p_1 + p_2 = 1$, are set by the quantum mechanism for Monday and Tuesday respectively, if the credences obey the two-day versions of the constraints $\widehat{S}_1^2$ and $\widehat{S}_2^2$ then the rational conditional degrees of belief for the two-day version are tabulated as follows:

	M	T
S	p_1	p_2
M	0	1

The only possible surprise day is Monday, and if Monday is to be a surprise day then p_1 must be less than or equal to p^*, the minimum required for justified belief, in which case the students are not justified in believing on Sunday that the exam will take place on a surprise day. In that sense, on the cut-off analysis of justified belief $\widehat{A}_4^2$ is a conditional blindspot.

7 Subsidiary paradoxes and puzzles

The cut-off analysis of justified belief has the virtue of illustrating some conceptual points and in showing how the mystery of the surprise exam paradox can be drained. But it is now time to consider some unattractive features of the analysis.

7.1 The lottery paradox

As already noted, the cut-off analysis violates the conjunction principle $JB1$: the conjunction of $P(X/\mathcal{E}) \geq p^*$ and $P(Y/\mathcal{E}) \geq p^*$ does not imply $P(X\&Y/\mathcal{E}) \geq p^*$. The failure of this implication is at the heart of the lottery paradox which Henry Kyburg (1961) framed in terms of acceptance rather than justified belief. Here are three candidate principles governing rational acceptance of a proposition:

> $RA1$: If it is rational for an agent to accept X and and rational for her to accept Y then it is rational for her to accept $X\&Y$.
>
> $RA2$: It is rational for an agent to accept X if her rational degree of belief in X conditional on the evidence available to her is sufficiently close to 1.
>
> $RA3$: It is not rational to accept a self-contradictory X.

To show that $RA1 - RA3$ are in conflict consider the set up exploited in the previous section: a lottery with N tickets with one ticket to be chosen the winner by means of a randomizing device that gives equal weight to each ticket. Then for an agent cognizant of how the lottery operates the rational degree of belief that any given ticket is *not* the winner is $1 - \dfrac{1}{N}$, and this degree of belief can be made as close to 1 as desired by choosing N sufficiently large. Thus, whatever cut-off value is set in $RA2$ (as long as it is strictly less than 1), the agent is rational in accepting the proposition $X(n)$ that ticket $\#n$ is a loser for every $1 \leq n \leq N$. By $RA1$ the agent is rational in accepting the conjunction $\&_{1 \leq n \leq N} X(n)$. But this conjunction is a contradiction (given the background information), violating $RA3$.

The lottery paradox can be resolved by rejecting the cut-off condition $RA2$ for rational acceptance or, more radically, by giving up on the idea of rational acceptance tout court in favor of degree of acceptance. Kyburg did not take either of these options but instead rejected the conjunction principle $RA1$ for acceptance.

7.2 Justified belief and the cut-off analysis

A number of commentators have wanted to retain the conjunction principle $JB1$ for justified belief (for example, see Ryan 1996, Evine 1999, Nelkin 2000, and Douven 2002), which obliges them to reject a probability cut-off analysis of justification—high rational degree of belief may be necessary for justified belief; but if the cut-off for degree of belief is set short of 1 by the tiniest epsilon it cannot be deemed sufficient without generating inconsistency. [Aside: For a (countably) infinite lottery setting the probability cut-off to 1 does not avoid the problem, at least not if the probability measure is merely finitely additive rather than countably additive. For there are mere finitely additive probability measures on the positive integers that assign probability 0 to each of the integers. If such a measure is generated by the mechanism for choosing the winning ticket then the probability is 1 that ticket $\#n$ is a loser for each $0 < n \leq \infty$. However, there is reason to think that such a mechanism cannot be realized by quantum mechanics. Most physicists assume, either explicitly or implicitly, that physically realizable quantum states are normal, and such states generate countably additive probability measures. Some of the reasons in support of this assumption are discussed in Earman (2020).]

Suppose, however, that $JB1$ is rejected, as it must be on the cut-off analysis of justified belief, mirroring Kyburg's rejection of the conjunction principle $RA1$ for rational acceptance. It is then not outright inconsistent for an agent to assert of each and every lottery ticket that she is justified in believing that it is a loser *and* also that she is justified in believing that some ticket is a winner. But the dual assertion has a distinctly odd ring, and for an obvious reason: the speaker is revealing that she is operating with a notion of justification which guarantees that, in certain situations, she will deem herself to be justified in believing a false proposition. This is so for the probability cut-off analysis of justified belief in the lottery case. Given any cut-off value $p^* < 1$, the number N of lottery tickets can be chosen so that $P(X(n)/\mathcal{E}) \geq p^*$ for each $1 \leq n \leq N$ and, thus on the cut-off analysis, $JB[\mathcal{E}](X(n))$ for each $1 \leq n \leq N$. And whatever the number N of lottery tickets and whatever the cut-off value p^*, $P(\exists_{1 \leq n \leq N} \neg X(n)/\mathcal{E}) = 1$ so that $JB[\mathcal{E}](\exists_{1 \leq n \leq N} \neg X(n))$ on the cut-off analysis. This feature does not automatically disqualify the cut-off analysis—that judgement depends on epistemic goals and a comparison of how well the cut-off analysis vs. other accounts of justification serve those goals (see Douven 2008); but it does seem that the proponent of the cut-off analysis needs to do some special pleading.

These considerations make me inclined either to reject a probability cut-off analysis for justified belief (and for acceptance as well) or else to abandon the unqualified notion of justified belief (and acceptance) in favor of degree of justification (and degree of acceptance). Additionally, the lottery example reinforced by the famous Gettier examples (see Gettier 1963) indicate that the probability cut-off analysis does not produce the kind of justification needed to sustain a knowledge claim. In the lottery example the high probability that a given ticket is a loser does not sustain the holder's claim to know that her ticket is a loser—or so many commentators would judge.[24] If you are sick of the lottery example consider Willard who uses the probability cut-off analysis of justification. He believes that X, and he proclaims that his belief is justified because given his available evidence $\mathcal{E}$ his rational degree of belief in X, $P(X/\mathcal{E})$, is very high, say, 0.999. (Willard's rational credence need not be derived from the statistical probability of a random lottery drawing; it might instead derive from the evidence of Willard's own perceptions and testimony of eyewitnesses.) Willard does not believe that Y, which is incompatible with X; indeed, Willard's $P(Y/\mathcal{E})$ is very low, say, 0.0001. It is a theorem of probability that if $P(X/\mathcal{E}) \geq p$ then $P(X \vee Y/\mathcal{E}) \geq p$, and so Willard proclaims that he is justified in believing that $X \vee Y$. Now suppose that $X \vee Y$ is true so that Willard has justified true belief according to the cut–off analysis of justification. All of this is compatible with supposing that $X \vee Y$ is true because X is false and Y is true. Under these conditions can we credit Willard's claim to know that $X \vee Y$ is true because he has justified true belief? Arguably not since his justification for believing $X \vee Y$ flows through through the false X which (on the cut-off analysis) he is justified in believing rather than through the true $\neg Y$ which he is also justified in believing.[25]

These examples also cast doubt on the wisdom of framing the surprise exam paradox in terms of knowledge or justified belief. This framing implies that the criterion of surprise is that on the evening before exam the students do not know or are not justified in believing (in a sense that supports a knowledge claim) that the exam will be on the morrow. What the above examples show is that by this criterion being surprised that X is true is

[24] Olin (1983) takes the surprise exam paradox and the lottery paradox to show that strong support by the available total evidence is not sufficient for justified belief. In another context Williamson (1996) holds that probabilistic evidence for X does not warrant asserting that X but only asserting that X is probable; see section 8.1 below.

[25] Sometimes the moral drawn from the Gettier examples is that knowledge is not just justified true belief—some fourth element is required. Other times the moral is said to be that the kind of justification required to maintain the knowledge-as-justified-true-belief slogan has to be more than merely having strong evidence in favor of the proposition—in particular strong evidence in the form of high rational degree of belief.

compatible with having a high rational degree of belief that X is true—the opposite of being surprised that X in any ordinary sense.

Is there a plausible way of construing the surprise exam announcement that does not entangle the resolution of the paradox with the unsettled wrangles of what constitutes knowledge and justified belief?

8 A student friendly version of the surprise exam announcement

When the students say they were surprised by the exam day the most plausible construal of their pronouncement is not that they didn't know or were not justified in believing that the exam would be on the morrow—at least not in a sense of knowledge or justification that generates debate in the philosophical literature. They simply mean that the exam was unexpected. If pressed to make "unexpected" more precise, they would probably agree that this means that on the night before the exam their rational degree of belief that the exam would be given the following day was low. How low? There is a grey area here, but if the value of the degree of belief is pushed sufficiently low the students will agree that they were surprised. Let v_1 be such a value (the precise numerical value won't matter to what follows). This notion of surprise satisfies a conjunction principle: taking $Sur[\mathcal{E}](\bullet)$ to mean that $P(\bullet/\mathcal{E}) < v_1$, if $Sur[\mathcal{E}](X)$ and $Sur[\mathcal{E}](Y)$ then $Sur[\mathcal{E}](X\&Y)$ since if $P(X/\mathcal{E}) < v_1$ and $P(X/\mathcal{E}) < v_1$ then $P(X\&Y/\mathcal{E}) < v_1$. And there is nothing untowards about asserting, before the drawing in a lottery, of each and every lottery ticket that one would be surprised if that ticket is the winner; indeed, to proclaim otherwise would be irrational.

If no exam occurs on days $1, ..., j$ and the students' memories are not failing they may say on the evening of day j that they know or are justified in believing that an exam has not taken place so far. But I doubt that they mean this in any sense of knowledge or justified belief that excites philosophical controversy. Rather, they simply mean that they have high confidence that no exam has taken place. If pressed further to explain "high" they would probably agree that they mean their credence is above $\frac{1}{2}$. How much above? There is a grey area here, but if the value of the degree of belief is pushed sufficiently high the students will agree that at this level they are confident. Let v_2 be such a value (the precise numerical value will not matter in what follows). As with cut-off analysis of justified belief, this notion of confidence does not satisfy a conjunction principle (if $v_2 < 1$): $Con[\mathcal{E}](Y)$ and $Con[\mathcal{E}](Z)$ does not imply $Con[\mathcal{E}](Y\&Z)$ since $P(Y/\mathcal{E}) \geq v_2$ and $P(Z/\mathcal{E}) \geq v_2$ does not imply $P(Y\&Z/\mathcal{E}) \geq v_2$. But there is no awkwardness about this failure of the conjunction principle for Con. In the lottery case, for example, there is nothing untowards about proclaiming before the drawing one that is confident of each and every ticket that it will not be the winner; indeed, in a lottery with a sufficiently large number of tickets to proclaim otherwise would be irrational. Whatever the exact values of v_1 set for Sur and v_2 for Con, as long as $v_2 > v_1$, $Con[\mathcal{E}](X)$ implies $\neg Sur[\mathcal{E}](X)$.

Using Sur and Con a student friendly formulation of a N-day surprise examination would have the form

$$A_5^N : (m \in \{1, ..., N\}) \& \forall_{0 \leq j \leq N-1}[(m = j + 1) \rightarrow Sur[\mathcal{E}_j](m = j + 1)]$$

and the N-day analogs of the supplementary conditions are

$$\widetilde{S}_1^N : (m > j > 1) \rightarrow Con[\mathcal{E}_j](m \notin \{1, ..., j\})$$

$$\widetilde{S}_2^N : Con[\mathcal{E}_0](m \in \{1, 2, ..., N\})$$

A surprise day i, where $1 \leq i \leq N$, is now defined as a day such that $P(m = i/\mathcal{E}_{i-1}) < v_1$ (the cut-off value for surprise).

Suppose that the exam day is chosen by a random mechanism that gives equal weight to each of the N days and that the students' degrees evolve per the sort of schedule laid out in Section 5. Then (i) $\widetilde{S}_1^N$ is satisfied because for $m > j > 0$, $P(m \notin \{1, ..., j\}/\mathcal{E}_j) = 0$; and whatever the choice of v_1 for Sur and v_2 for Con, there is a sufficiently large N such that A_5^N is true if the random mechanism chooses a surprise day for the exam. This is all compatible with satisfying $\widetilde{S}_2^N$ by making $P((m \in \{1, 2, 3, 4, 5\}))/\mathcal{E}_0) = 1$. And (ii) on Sunday the students can be confident the exam will fall on a surprise day. That confidence will wain as the week progress without an exam. But that confidence loss engenders no paradox; indeed, this waning confidence is what avoids paradox since it goes hand in hand with the rise in the expectation that the exam will occur the following day. When that expectation rises to the level of v_1 the backwards induction is halted.

8.1 The price the teacher pays for setting a surprise exam

Using the A_5^N-$\widetilde{S}_1^N$-$\widetilde{S}_2^N$ version of the surprise exam announcement with sufficiently large N the teacher has succeeded in setting a surprise exam by using a random drawing to set the exam day and by revealing her method in advance to the students. But in doing so she has, arguably, committed a sin: her announcement—construed as an assertion—violates a constitutive norm of assertability. Williamson (1996)[26] provides powerful reasons for the *knowledge rule* of assertability:

> One must: assert that X only if one knows that X.

Combining the knowledge rule with the view, endorsed above, that mere high probability resulting from a random draw does not sustain a knowledge claim makes the teacher's sin manifest. If the exam falls on day N then $A_5^N \& \widetilde{S}_1^N$ is false; this is statistically unlikely, but on the endorsed view of knowledge the teacher does not know when she makes her announcement that the exam will not fall on day N no matter how large N and, thus, no matter how low the probability of an exam on day N. Nor, according to Williamson, is the teacher's sin expiated by replacing the knowledge rule by the *warrant rule* of assertability:

> One must: assert that X only if one has warrant to assert X.

For on Williamson's view, "[p]robabilistic evidence warrants only an assertion that something is probable" (Williamson 1996, p. 501).

The precocious students begin to wonder whether their teacher has intentionally violated norms of assertability in order to inflict a surprise exam on them. Their doubts fester, leading to suspicions about what other norms of engagement the teacher may have violated and poisoning the atmosphere of the class. Because of the bad teaching reviews she receives the teacher is not promoted. She finds work as a Starbucks barista which, to her surprise, she finds more rewarding than teaching philosophy. Sadly, she is laid off during the Covid pandemic and is now on welfare. Such are the wages of tangling with the surprise exam paradox.

[26] I am grateful to David Baker for calling this reference to my attention.

9 Conclusion

Since the surprise exam announcement is open to so many different readings it is hardly surprising that, despite all the effort philosophers have devoted to it, there is no definitive resolution to be found for "the surprise exam paradox." But has all the effort been rewarded by commensurate philosophical insights that would mark the surprise exam as a worthy paradox? No, or at least nothing in the present review has uncovered any insight that couldn't have been gleaned from other studies. The claim of the Introduction seems vindicated: the problem with the surprise exam paradox lies not in resolving the paradox(es) but in pinning down a worthy paradox.

Zombie like, the surprise exam lives on, refusing to die. There are two reasons, one meretricious the other admirable. The former flows from the fact that philosophers have many hammers in their tool kit and these hammers need to find nails to pound, even if the pounding results largely in a lot of noise. The more laudable reason is that the paradox reveals itself to be a wonderful teaching instrument. A class demonstration of, say, the Sorensen spatialized version of the paradox can be used to engage students' interest, and once the students are hooked, their engagement with the paradox can be used to motivate them explore issues in logic and epistemology.

Acknowledgmnent: I am indebted to David Baker, Gordon Belot, Anil Gupta, John Norton, and Laura Ruetsche for helpful comments on drafts of this paper. But, needless to say, this does not imply that they agree with any of the opinions expressed herein.

References

[1] Ardeshir, M. and Ramezanian 2012. "A Solution to the Surprise Exam Paradox in Constructive Mathematics," *Review of Symbolic Logic* 5: 679-686.

[2] Auman, R. T."Backward Induction and Common Knowledge of Rationality," *Games and Economic Behavior*," 8: 6-19.

[3] Ayer, A. J. 1973. "On a Supposed Antinomy," *Mind* 82: 125-126.

[4] Binkley, R. 1968. "The Surprise Examination in Modal Logic," *Journal of Philosophy* 65: 127-136.

[5] Bovens, L. 1997. "The backwards induction argument for the finite prisoner's dilemma and the surprise exam paradox," *Analysis* 57.3: 179-186.

[6] Chihara, C. 1985. "Olin, Quine, and the Surprise Examination," *Philosophical Studies* 47: 191-199.

[7] Chow, T. Y. 1998. "The Surprise Examination or Unexpected Hanging Paradox," *American Mathematical Monthly* 105 (1998): 41-51.

[8] Chow, T. Y. 2011. "The Surprise Examination or Unexpected Hanging Paradox," URL arXiv:math/9903160.

[9] Douven, I. 2002. "A New Solution to the Paradoxes of Rational Acceptability," *British Journal for the Philosophy of Science* 53: 391-410.

[10] Douven, I. 2008. "The Lottery Paradox and Our Epistemic Goal," *Pacific Philosophical Quarterly* 89: 204-225.

[11] Earman, J. 2020. "Quantum Sidelights on The Material Theory of Induction," *Studies in History and Philosophy of Science Part A* 82: 9-16.

[12] Evine, S. 1999. "Believing Conjunctions," *Synthese* 118: 201-227.

[13] Fitch, F. B. 1964. "A Gödelized Formulation of the Prediction Paradox," *American Philosophical Quarterly* 1: 161-164.

[14] Gardner, M. 1963. "A new paradox, and variations on it, about a man condemned to be hanged. *Scientific American* 208 (March):144–154.

[15] Gerbrandy, J. 2007. "The Surprise Examination Dynamic Epistemic Logic," *Synthese* 155: 21-33.

[16] Gettier, E. L. 1963. "Is Justified True Belief Knowledge?" *Analysis* 23 (6): 121-123.

[17] Hall, N. 1999. "How to Set a Surprise Exam," *Mind* 108: 647-302.

[18] Halpern, J. Y. and Moses, Y. 1986. "The Paradox of the Surprise Test Revisited," *Journal of Philosophical Logic* 15 (1986): 281-304.

[19] Harrison, C. 1969. "The Unanticipated Examination in View of Kripke's Semantics for Modal Logic," in Davis J. W., Hockney D. J., Wilson W. K. (eds) *Philosophical Logic*, pp. 74-88. Synthese Library, vol 20. Springer: Dordrecht.

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

[21] Kritchman, S. and Raz, R. 2010. "The Surprise Examination Paradox and the Second Incompleteness Theorem," *Notices of the American Mathematical Society* 57: 1454-1458.

[22] Kyburg, H. 1961. *Probability and the Logic of Rational Belief*. Middletown, CT: Wesleyan University Press.

[23] Lewis, D. 1980. "A Subjectivist Guide to Objective Chance." In R. Carnap and R. Jeffrey (eds.), *Studies in Inductive Logic and Probability*, pp. 263-293. Berkeley: University of California Press.

[24] McLelland, J. and Chihara, C. 1975. "The Surprise Examination Paradox," *Journal of Philosophical Logic* 4: 71-89.

[25] Nelkin, D. 2000. "The Lottery Paradox, Knowledge, and Rationality," *Philosophical Review* 109: 373-409.

[26] O'Connor, J. 1948. "Pragmatic Paradoxes," *Mind* 57: 358-359.

[27] Olin, D. 1983. "The Prediction Paradox Resolved," *Philosophical Studies* 44: 228-233.

[28] Quine, W. V. 1953. "On a So-Called Paradox," *Mind* 62: 65-67.

[29] Raatikainen, P. 2020. "Gödel's Incompleteness Theorems," *The Stanford Encyclopedia of Philosophy* https://plato.stanford.edu/archives/win2020/entries/goedel-incompleteness/.

[30] Ryan, S. 1996. "The Epistemic Virtues of Consistency," *Synthese* 109: 121-141.

[31] Scriven M. 1951. "Paradoxical Announcements," *Mind* 60 (1951): 403-407.

[32] Shaw, R. 1958. "The Paradox of the Unexpected Examination," *Mind* 67: 382-384.

[33] Sobel, J. H. 1993. "Backwards induction arguments: paradox regained," *Philosophy of Science* 60: 133-144.

[34] Sober, E. '998. "To Set a Surprise Exam Use Game Theory," *Synthese* 115: 355-373.

[35] Sorensen, R. 1982. "Recalcitrant Versions of the Prediction Paradox," *Australasian Journal of Philosophy* 60: 355-362.

[36] Sorensen, R. 1984. "Conditional Blindspots and Knowledge Squeeze: A Solution to the Prediction Paradox," *Australasian Journal of Philosophy* 62: 125-136.

[37] Sorensen, R. 1988. *Blindspots.* Oxford: Oxford University Press.

[38] Sorensen, R. 2017. "Epistemic Paradoxes," *Stanford Encyclopedia of Philosophy URL* https://plato.stanford.edu/archives/fall2020/entries/epistemic-paradoxes/.

[39] Williamson, T. 1996. *Philosophical Review* 105: 489-523.

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