Practical Solutions to the Surprise Examination Paradox

I

Can a teacher who promises to give a surprise examination during the following week keep his word? The reasoning purporting to show that the promise cannot be fulfilled is familiar. The exam cannot take place on the final day, because it will not be a surprise. Having eliminated the final day, the penultimate day can be analogously eliminated, and so on, until the first day is ruled out. On no day will the exam be surprising, and the promise cannot be kept.

If a paradox is "an apparently unacceptable conclusion derived by apparently acceptable reasoning from apparently acceptable premises",[1]then to solve it, we must either embrace the conclusion (and make it intellectually palatable), find a fault in the derivation, or reject a premise in a principled way.

Some paradoxical conclusions cannot be accepted. In the racetrack paradox Zeno argues that in order to get from A to B one must get to the midpoint. That requires that one, first, get to the point midway between A and the midpoint, and so on, *ad infinitum*. Since one cannot complete infinitely many journeys, one cannot get from A to B. But these are two arbitrarily chosen points, so motion is impossible.

What is the extent of the absurdity? Well, Zeno's reasoning cannot be applied to appearances, because visual space isn't infinitely divisible. So it is just the possibility of motion in objective, external reality that is being denied. But even so, that means our senses systematically deceive us about motion, and casts doubt on their reliability regarding other matters (objects' shapes, for instance), as well. Pending an account of how static reality affects our senses, so as to engender deceptive appearances of motion - in analogy with the explanation we now give of

---

[1] See M.D. Sainsbury, *Paradoxes* (Cambridge: Cambridge University Press, 1988.), p. 1.

optical illusions, the paradox shows that we are incapable of transcending the "veil of appearances".

The sense of absurdity engendered by a paradoxical conclusion can sometimes be dispelled. We admit that there can be no barber who shaves precisely those who do not shave themselves. Our initial error is easily accounted for, without engendering general doubt about our beliefs' reliability: logical intuitions are fallible, and our (naive) opinion about the barber is just one case where they lead us astray. The detection of the error does not cast doubt on our ability to reason. Indeed, it is the result of reasoning rigorously. (This is different in the case of Zeno, who uses one putative source of knowledge - deduction - to undermine *another* - the senses.)

II

Is the surprise examination paradox akin to Zeno's? Why has it seemed (almost without exception) obvious that something "*must* be wrong with the way in which the class reasoned"?[2] Well, the reasoning must be fallacious if it purports to show that the promise is inconsistent. For the announcement can be made true: if the teacher gives an exam on Tuesday (say), the students *will* be surprised. And as Quine[3] has pointed out, even upon hearing the announcement, the students cannot rule out the logical possibility that they will be surprised if an exam is given on the final day. Hearing the announcement provides no logical guarantee that the students will expect an exam on the final day if one hasn't yet been given. So the promise is not *contradictory*.

[2] Sainsbury, *Paradoxes*, p. 94, my italics.
[3] See W.V.O. Quine, 'On a So-called Paradox', *Mind* 62 (1953).

This version of the paradox deserves, perhaps, Quine's disparagement. But the paradox is more resilient. To be sure, the promise can (*logically*) be kept. The students may not understand it; they may subsequently forget it; or they may conclude by *reductio* that the announcement gives them no rational grounds for *any* expectation. But these *outré* possibilities are insufficient to render the promise *credible*. And on its credibility depends the best (perhaps the only) explanation of a familiar (classroom) phenomenon. We attribute to the promise a *communicative role*. The announcement conveys information, which the students believe (knowing the teacher to be reliable). This belief motivates them to behave in a certain way, a fact which the teacher utilises - in making the promise - to promote some aim.

More specifically, the teacher would like the students to study daily, and knows that the more likely a student thinks the teacher is to give an examination the following day, the more he will rehearse the material (if he is sufficiently studious). Given this aim, the teacher would do best to give daily exams (and publicise his intention). But he cannot give more than one, because he must leave ample time for teaching. Keeping the students in a state of suspense about the actual day is his optimal strategy.

The cogency of the explanation is threatened by the students' reasoning. For how can an unbelievable (albeit consistent) promise have the effect that it does on its hearers? To function in the classroom, the promise must be *credible*, not just consistent, and the fault in the derivation must be found. That is why, as Williamson says,[4] any "adequate diagnosis of the Surprise Examination should allow the pupils to know that there will be a surprise examination".

[4] See T. Williamson, 'Inexact Knowledge', *Mind* 101 (1992), p. 230.

There is an important insight here. It enjoins us to adopt a *practical* stance towards the paradox; to view the promise as a real life utterance, whose role a solution must reflect. This is not to say that the paradox poses a practical problem. Rather, a solution to what is a *theoretical* problem must incorporate an explanation of a certain *practice*. Let us consider how this conception determines what is to count as an adequate solution to the paradox.

III

If we approach the paradox practically, we will rule out Quine's and Olin's[5] solutions. They explain that the students can *be* surprised because they ought not to believe the promise. Rather than providing a solution to the practical paradox, they reinforce it.

Second, from a practical point of view the sting is taken out of Sorensen's observation that some solutions to the surprise examination paradox invite formulations of more resistant strains of the paradox.[6] One such solution is offered by McLelland and Chihara.[7] They suggest that in deriving the paradoxical conclusion, the students rely on a principle of knowledge, $Kp \rightarrow KKp$, which isn't valid. So, they argue, the students' reasoning isn't cogent. In response, Sorensen offers the following version of the paradox. Five students are lined up, so that each can see the backs of those in front of him. One student has a star on his back. The teacher announces that the designated student is to be examined and will not know this till the students break

---

5 See D. Olin, 'The Prediction Paradox Resolved', *Philosophical Studies* 44 (1983).

6 See R. Sorensen, 'Recalcitrant Variations of the Prediction Paradox', *Australian Journal of Philosphy* 60 (1982).

7 See J. McLelland and C. Chihara, 'The surprise Examination Paradox', *Journal of Philosophical Logic* 4 (1975).

formation. Sorensen points out that unlike the original promise, *this* can be shown to be paradoxical without invoking the principle McLelland and Chihara reject. So here the reasoning *is* cogent.

Does Sorensen's observation undermine McLelland and Chihara's solution to the paradox? Is it true, as Sorenson says,[8] that "an adequate solution . . . must be a solution for *all* of the variations"? Not if we view the paradox *practically*. Since this new version of the promise is never uttered (let alone believed) *in real life situations*, we can easily resist the pressure to render it consistent. Here, we may concede the conclusion by way of solving the paradox. Perhaps in some artificial settings we can derive a paradoxical conclusion without the epistemic principle Kp→KKp. Perhaps we cannot coherently make the designated-student promise. But that tells us nothing about the *classroom promise*, since *it* is paradoxical only in conjunction with the epistemic principle.

I am not endorsing the solution offered by McLelland and Chihara: it crucially depends on the rejection of a *prima facie* plausible epistemic principle. My point is, rather, that the objection levelled against it by Sorensen is misguided from a practical perspective, constrained by the need to explain how seemingly paradoxical utterances *actually* function.

Practical considerations can, thirdly, be invoked against a solution to the paradox, proposed by Wright and Sudbury.[9] The backward induction, they suggest, cannot get underway without the assumption that the promise will continue to be credible on the last day even if the exam has not been given by then. Without this premise (the "temporal retention of knowledge principle"), the final day cannot be

---

[8] See R. Sorensen, 'Conditional Blindspots and the Knowledge Squeeze', *Australian Journal of Philosophy* 62 (1984). The italics are mine.

[9] See C. Wright and A. Sudbury, 'The Paradox of the Unexpected Examination', <u>Australian Journal of Philosophy</u> 55 (1977).

ruled out. The exam will surprise the students even then, because they will no longer believe the announcement. In Sorensen's words,[10] it will have become an "epistemological *blindspot*": a consistent proposition that cannot be believed. (Another example of a blindspot is provided by the proposition 'p and I will never find out that p'.)

This solution is ingenious, but the "offending" premise cannot just be denied. Unless the students believe that the promise will lose its credibility if it hasn't been kept before the final day, we do not have a practical solution. But can we impute this belief to ordinary students? Promises are customarily assumed to remain credible. Of course, once we realise that the paradoxical promise can be kept only if it ceases (at some point) to be credible, we may withdraw the assumption. But we cannot suppose that the students have done so. Since *everyone* is surprised by the derivation, ordinary students, who are not familiar with the paradox, have had no reason to withdraw a presumption (implicitly) applicable to promises.

IV

On the positive side, two solutions can be motivated by the practical conception of the paradox. First, once the promise is viewed as a real life utterance, the suggestion arises that it might be construed *non-literally*, as are some other utterances. For instance, might not the teacher be interpreted as promising an exam *sometime* during the following week? (Indeed, this is what many teachers actually promise.) This interpretation, admittedly, weakens the teacher's commitment to surprise the students. For instance, if he gives the exam on a randomly chosen day, he

[10] Sorensen, 'Conditional Blindspots and the Knowledge Squeeze'.

can guarantee that the exam will be given, but there is a small probability (1/5) that it will not be surprising. But there is a reason for preferring the non-literal construal of the promise. It seems to reflect the teacher's purpose (as described above) better. His aim isn't to catch the students off guard; the surprise is just a means to an end. And it isn't in his interest that the exam should be surprising if it is given on the final day. If the non-literal reading is adopted, the students will be maximally motivated to study in those circumstances, because they will expect an exam with certainty.

Note that the non-literal interpretation of an utterance is based on background knowledge. The teacher's intentions, for instance, may favour the non-literal construal of his promise cited above. But this suggests that utterances with the same formal structure, "There will be a surprising event of type E during the following n days", may invite *different* non-literal interpretations. And this is, indeed, what we find. The paradoxical statement 'There will be a surprising earthquake in San Francisco sometime during the next 100 years' is best interpreted by imputing to the speaker a *full* commitment to the earthquake being surprising, and weakening his guarantee that it will take place within the specified period. (There is a tiny probability that the earthquake will not occur.)

V

So much for the first practically-motivated solution. The second such solution invites us to go back and examine more closely the assumption that in order to solve the paradox we must find a fault in the reasoning. Of course, to be effective, the promise must be *credible*. And of course, a statement which is believed to be paradoxical will not be credible. And, furthermore, ordinarily, statements are credible because they *are* non-paradoxical. The most straightforward way of solving the paradox, then, is to show that the promise is - initial appearances to the contrary - non-paradoxical. But there is an alternative.

A statement may be very subtly paradoxical, or even contradictory. And if it is mistaken for a non-paradoxical one, as in the case of the barber, it may be reasonably believed. Logical omniscience isn't a necessary condition for rationality. When this is appreciated, the pressing need to find the fallacy in the students' reasoning is removed. It is perfectly reasonable for ordinary students to believe the promise even if the reasoning which shows it to be paradoxical is impeccable. They know the teacher is typically credible, and do not realise *this* announcement wouldn't be believed by a logical wizard: reasoning by backward induction does not come very naturally.

If "the student's argument is not to be open to trivial objections", Olin argues,[11] he must be assumed to be "an expert logician' so that if he were justified in believing $p_1,...p_n$, which jointly imply . . . q, then he would see that $p_1,...p_n$, jointly imply . . . q". Not so. Olin dismisses the objection that students aren't logically perfect as "trivial". But her idealisation takes us away from the real life situation which called for an explanation. Indeed, it impugns a cogent explanation.

The practical solution which appeals to the difference between *perceived* and *actual* paradoxicality enables us to account for an aspect of the real life phenomenon: the difference between the one-day and the five-day version of the paradox. The announcement 'There will be a surprise exam tomorrow' is palpably paradoxical. No great logical acumen is required to perceive this, and that is why the announcement is incredible even for ordinary students, who do not spot the (subtler) paradoxicality of

[11] Olin, 'The Prediction Paradox Resolved', p. 227.

the statement ‘There will be a surprise examination next week’. Of course, other solutions (with the exception of Quine’s and Olin’s) can account for this datum, as well. According to McLelland and Chihara,[12] Wright and Sudbury,[13] and Sorensen,[14] only the one-day version is paradoxical. According to Williamson,[15] the one-day case is definitely paradoxical, but an n-day (n>1) version might be paradoxical for one person without being paradoxical for another. The students’ reasoning invokes the temporal retention principle: I know on the first morning that I will know on the second morning that…I will know on the last morning the truth of the teacher’s announcement. And, Williamson claims, the number of iterations we are legitimately allowed to invoke isn’t uniform: some people have better knowledge of their future knowledge.

There remains the question of how we are to *interpret* the paradoxical promise as understood by the students. It is not enough to suppose that they do not realise it is paradoxical. The statement *is* paradoxical - at least when conjoined with plausible assumptions (Kp→KKP, e.g., or the knowledge retention principle). And, as Wright and Sudbury point out,[16] a solution “should make it possible for the pupils to be *informed* by the announcement: we want the reaction of someone who notices no peculiarity but just gets on with his revision to be logically unobjectionable”. Now, there is no difficulty in accounting for our reaction to a non-paradoxical announcement. If the teacher promises an exam tomorrow, we just go and revise. And we might even suppose that we can harbour contradictory beliefs, so long as the inconsistency isn’t flagrant, and remains undetected and *dormant*, so to speak. But

[12] McLelland and Chihara, ‘The surprise Examination Paradox’.
[13] Wright and Sudbury, ‘The Paradox of the Unexpected Examination’.
[14] Sorensen, ‘Conditional Blindspots and the Knowledge Squeeze’.
[15] Williamson, ‘Inexact Knowledge’, p. 231.
[16] Wright and Sudbury, ‘The Paradox of the Unexpected Examination’, p. 42, original italics.

how does one *act* on a contradictory (paradoxical) statement? Again, there is here pressure to render the statement consistent.

We should resist the pressure. Without providing a general dispositional account of inconsistent beliefs, we can say something about the paradoxical promise and the behaviour it engenders. The students fully expect a surprise exam to be given during the week. They assign a non-negligible probability to it being given on the first day, and on other days of the week, and that is why they will study during the week until the exam is given - so long as the teacher doesn't wait too long, allowing them to realise (as they might - come Thursday evening) that the promise was, in fact, paradoxical.

VI

It might be objected that even if we understand how ordinary students can believe the promise and respond appropriately, we face a difficulty when we try to account for the effect it might have on sophisticated students, who have gone through the reasoning. For them, surely, the statement is incredible!

In response, we should distinguish two cases. If the teacher is assumed to be naive, then the students should simply respond to the promise as does a naive student. By the end of the week, if the teacher hasn't given an exam, he will himself realise he has made a paradoxical promise. And it will not be clear what to expect he will do on Friday. But this, note, is not part of the content of the original promise. When it was given, the teacher fully expected his promise to remain credible.

The second case to consider is one in which the paradox becomes common knowledge: the students are familiar with it; the teacher knows they are; the students know he does, and so on *ad infinitum*. Thus, a philosophy lecturer, having explained the paradox to a class, might promise to give a surprise examination during the remainder of term. How are the students to react? What are they to make of the announcement? If it is not simply made as a joke, they will do best to construe the statement in accordance with the lecturer's preferred solution. Thus, they may suppose that he intends his promise to lose its credibility by the end of the week.

Note how the practical conception of the paradox is at work here. There is no *right* way of construing the statement *in general*: we are trying to decide the intention of the *speaker*, and here, sophisticated speakers may greatly differ. Note also that the non-paradoxical interpretation of the announcement is only made plausible by the students' knowledge that the lecturer knows there is a consistent interpretation, and has it in mind when he makes his utterance. If he thinks (following Quine and Olin) that the promise is incredible, *they* should so treat it, and view the announcement as uninformative.

The above consideration takes us even further away from a *general* solution to the paradox. I cautioned above (section III) against the assumption that all the versions of the surprise examination paradox have the same solution. We now see that we might have to interpret differently the same classroom promise, depending on what we know about its utterer. Of course, what motivates this proliferation of solutions is one conception of the paradox: the practical one.