Knowing about Surprises: A Supposed Antinomy Revisited

CHRISTOPHER JANAWAY

A given event may be a surprise to you, even if you know that it is going to occur. It may be a surprise to you, even if you know that it is going to occur and be a surprise to you. But what is not possible is that you should know a finite list of possible times at which it may possibly occur, and know that it will be a surprise to you. I shall argue that this is sufficient to dispense with a well-known paradox or antinomy.

I

The problem may be stated like this: A student S is informed by a teacher that he will be given a test on one of a series of occasions t_1–t_n, and that he will be surprised when the test occurs. He reasons, however, that t_n is not really one of the possible occasions on which the test may occur, since he already knows that, once t_1–t_{n-1} have elapsed, the test cannot be a surprise to him. But if the series of possible occasions is assumed to be t_1–t_{n-1}, then, by the same argument as before, he reasons that the test will not occur at t_{n-1}. This reduces the series of possibilities to t_1–t_{n-2}, and so on. S concludes by this argument that it is not possible for the test to occur at a time when its occurrence will be a surprise to him—that is, at a time such that, *prior to* that time, he does not know that it will occur *at* that time. This is the argument for what I shall call the Thesis side of the antinomy, whose conclusion is that it is not possible for S to be surprised by the occurrence of the test. The argument for the Antithesis is simply of the form: p, so possibly p. The teacher gives the test at t_{n-2}, which S did not, prior to t_{n-2}, know she would do. So, it is possible for S to be surprised by the occurrence of the test.[1] In what follows I shall refer to the event concerned as E, and shall consider first the case where the possibilities are t_1–t_5.

[1] Credit for stating in print that this kind of example involves a serious antinomy goes to Michael Scriven, 'Paradoxical Announcements', *Mind*, 1951, pp. 403–7. D. J. O'Connor, in 'Pragmatic Paradoxes', *Mind*, 1948, pp. 358–9, had considered his example (the 'Class A blackout') to contain simply a 'pragmatically self-refuting' prediction, and to be in itself a 'rather frivolous example'. But as Scriven pointed out, *if* the prediction is self-refuting *and* can in reality be carried out, the matter is a little more serious: it is this 'flavour of logic refuted by the world' which 'makes the paradox rather fascinating'.

In his paper 'On a Supposed Antinomy',[2] Quine offers to reveal a fault in the argument for one half of the antinomy in question. Quine suggests that the reasoning for the Thesis is faulty. At the outset, that is before t_1, S looks ahead to what the situation will be between t_4 and t_5. Here, says Quine, S discerns only two alternatives: (*a*) E will have occurred at one of t_1–t_4, (*b*) E will occur at t_5, and S will know prior to t_5 that it will occur at t_5. Rejecting (*b*) because of its violation of the original decree, S elected (*a*), thus starting the process of whittling down the series until there remains no possible time at which E could occur and surprise him. Quine's objection is that there are not two, but four alternatives which S should have recognized, four ways in which matters could stand between t_4 and t_5. Apart from (*a*) and (*b*), S must reckon with: (*c*) E will fail to occur at t_5, and (*d*) E will occur at t_5, and S will not know prior to t_5 that it will occur at t_5. That is to say, S should have taken into account from the outset the possibility that the decree will be violated. He neither knows at the outset, nor will he know after t_1–t_4 have all elapsed, that E will occur at all. If it does fail to occur, he cannot have known it was going to occur. But given that there is the possibility of its failing to occur, then even if it does occur on the last occasion, he cannot have known that it would occur then.

Now A. J. Ayer[3] has suggested that there is a deeper fallacy in this example which Quine's proposal is inadequate to reveal. Quine, incidentally, substituted a judge and a surprise hanging for a teacher and a surprise test. What Ayer does in effect is have us substitute an amended prediction on the part of the judge. In Quine's version the prediction was:

> S will be executed at one of t_1–t_n, and will not know prior to the time of execution what the time of execution is.

But we can easily change this to:

> S will be executed at one of t_1–t_n, and will not know prior to the time of execution *that it will be the time of execution, if there is to be an execution at all.*

Quine must concede that S *can* know, when only t_n remains, *that* t_n *will be the time of execution, if there is to be an execution at all.* So by reasoning analogous to that above, S ought to be able to know from the beginning that the amended prediction will be false—which, as Ayers says, 'would still appear paradoxical'.[4]

I shall suggest below another way of showing that Quine's proposal is inadequate. But first let me comment on Ayer's own proposed solution.

[2] In *The Ways of Paradox*, pp. 21–3. The same paper was published under the title 'On a So-called Paradox' in *Mind*, 1953, pp. 65–7.

[3] 'On a Supposed Antinomy', *Mind*, 1973, pp. 125–6.

[4] Ibid., p. 125.

According to Ayer, a fallacy arises from arguing 'that because there could be circumstances in which all uncertainty had been removed [E's not having occurred at any of t_1–t_{n-1}], there is no uncertainty at the start'.[5] This, as far as it goes, seems to be right; but where precisely is the fallacy involved?

Ayer claims that a certain distinction is what 'solves the puzzle', namely that between knowing '*before the sequence is run through*' when E will occur, and knowing '*in the course of the run, however long it continues*' when E will occur. When he comes to apply this distinction to the puzzle, Ayer sets things up as follows: the planned execution may occur at one of only two times, t_1 and t_2, and if the prediction that S will be surprised when the planned time arrives turns out false, then S will be released. But Ayer thinks that the condition of S's release, and hence the sense in which the prediction can turn out false, is capable of two interpretations, which must be kept separate:

> If the condition of the prisoner's escaping [of the prediction's turning out false] is that he knows which day has been selected, he does not escape [the prediction does not turn out false], since he does not know, though he has an even chance of correctly guessing, which day it is. But if the condition is that there could be a time at which he would know which day he had been selected, he does escape [the prediction turns out false], since this time would come if the execution were set for the second day.

If we expand on this a little, we find that for Ayer there are two distinct possible predictions which the example might embody. These predictions are:

(A1) S will not be able to know, before any of t_1–t_n, when E will occur,
(A2) There is no time during t_1–t_n at which E could occur, such that prior to its occurring S will know when E will occur.

Now we can readily agree that (A1), *pace* the Thesis argument, is true, and that (A2) is clearly false. But what Ayer fails to explain is why there should be any temptation to confuse (A1) and (A2), or to infer the falsity of (A1) from the falsity of (A2). And it is therefore difficult to see in what way the distinction he draws our attention to does solve the puzzle.

However, there is a bigger drawback to Ayer's proposal. For the prediction in the example as traditionally discussed is in fact neither (A1) nor (A2). It is rather

[5] Ibid., p. 126. All further references to Ayer are to this same page.

S will not know prior to E's occurrence, that E will occur when it does.[6]

The condition of the prediction's turning out true was never that *it was impossible for it to be false*. Likewise, the condition of its failing to be true was supposedly not that S should know, *at the outset*, *when E will occur*, but rather that whenever E does occur, S should know of its occurrence prior to its occurrence. In fact, the falsity of (A2) is compatible with the truth of the prediction in the original example. And the truth of (A1) is compatible with the original paradoxical claim that the prediction cannot turn out true, and even with S's knowing 'before the sequence is run through' that the prediction cannot turn out true.

I conclude that Ayer's attempt to 'solve the puzzle' does not do so.

II

As I said earlier, there is another way of revealing that Quine's proposal is inadequate. And I hope to show that by pursuing it we can genuinely lay the antinomy to rest. The example as given by Quine admits of S's inability to know that the decreed event will occur; and it seems to be because he does not know that it will occur that it can occur in such a way as to surprise him. One reason why this is unsatisfying is that we can construct a variant case, of which it is true that S knows at the outset that E will occur, and hence knows at the outset that only Quine's alternatives (a) and (b) are open. If S knows at the outset that E will occur at just one of t_1–t_5, then S knows that if E has not occurred at any of t_1–t_4, E will occur at t_5.

Here is such a case. I put a coin into one of five indistinguishable boxes, B_1–B_5, to which only I possess a key. Without my seeing, someone else, who knows which box the coin is in, arranges the boxes in a line, and tells me to open them in turn at t_1–t_5, adding that I will not know, until I see the coin at a certain time, that I shall see it at that time. We can assume that in this case I know that there will be an event E of my seeing the coin, and that if I have not seen the coin at any of t_1–t_4, I shall know prior to t_5 that I will see it at t_5. So Quine's alternatives (c) and (d), which his condemned man allegedly erred in overlooking, are excluded from my reasoning *ex hypothesi*. After t_4 but before t_5, *either E* will have occurred, *or* I will know that E will occur at t_5, and there will be no surprise. If we are

[6] Earlier writers insisted correctly that E must be an event which S *cannot* or *could not* know in advance will occur when it does. Sometimes the stipulation is that E be an event which S *cannot know by deduction* will occur when it does (see O'Connor (1948), p. 358; Scriven (1951), p. 405; Martin Gardner, *Scientific American*, March 1963, p. 149). To make the formulation simpler, I simply take it for granted that the prediction would not be fulfilled by instances of S's 'not knowing' such as S's making an oversight or lacking the capacity to reason. More recently, as we shall see, some writers on the paradox have selected justified or reasonable belief as the notion in terms of which the paradox is best stated. Where the choice between one of these notions and knowledge seems to matter, I will indicate as such.

to take Quine's insertion of alternatives (c) and (d) as *the* way to block the argument for the Thesis, then we have our antinomy again. I should be able to reason that E will occur at just one of t_1–t_4, but that if it occurs at t_4 it will not surprise me, so that it must occur at one of t_1–t_3, and so on. The conclusion is the same: it is not possible for E to surprise me. Yet, it must be possible for it to surprise me, since (let us assume) I open B_3 and discover the coin without knowing that I was going to, and my friend's prediction comes true.

The solution cannot be, then, that S is not able to know that the event will occur. By stipulating a case which (it seems) differs from the original only in that S *does* know this, we resurrect the paradox.

III

This suggests that any solution that depends on the possibility of S's being surprised by E's occurring at t_n—the last available occasion—is too restricted. Quine's is one such suggestion, but many have objected that his particular answer makes the solution too easy:[7] we should assume that S is genuinely informed by the original announcement, and that S has good reason to believe it true. More recent writers who have made such stipulations, and who do not approve of Quine's proposed solution, have nevertheless followed Quine in arguing that S can be surprised by an occurrence of E on the last occasion provided for in the announcement.

Crispin Wright and Aidan Sudbury,[8] in a sophisticated account based on the notion of reasonable belief rather than knowledge, provide one such instance. Taking the original teacher example, they suggest that if t_{n-1} has passed and E has not occurred, then the reasonable course for S will be to suspend belief in *both* the proposition that E will occur at one of t_1–t_n and the proposition that E will be a surprise. This is of course to say that S will have no reason to believe the original announcement if t_{n-1} has elapsed and the test has not yet occurred. S will have no reason to believe any part of the announcement—since the two propositions mentioned make up the content of the announcement. Given that 'S is surprised at t' means 'S has no reason to believe prior to t that E will occur at t, and E occurs at t', then it is open to the teacher to surprise S on the last occasion.[9]

This has the same ultimate effect as Quine's proposal—namely that the Thesis argument cannot get off the ground because a surprise at t_n cannot be ruled out. Where Wright and Sudbury disagree with Quine is in their

[7] Cf. Charles S. Chihara, 'Olin, Quine, and the Surprise Examination', *Philosophical Studies*, 1985, p. 196; Crispin Wright and Aidan Sudbury, 'The Paradox of the Unexpected Examination', *Australasian Journal of Philosophy*, 1977, pp. 42–3; Roy A. Sorensen, 'Conditional Blindspots and the Knowledge Squeeze: A Solution to the Prediction Paradox', *Australasian Journal of Philosophy*, 1984, p. 129. [8] Wright and Sudbury (1977), pp. 41–58. [9] Cf. ibid., pp. 54, 55, 57.

insistence that S can have good reason to believe the teacher's announcement, in full, at the outset. The error in the Thesis argument, they claim, is that of assuming a principle of temporal retention of reasonable beliefs, whereby anything that it is reasonable for S to believe at a time t, it is reasonable for S to believe after t. If S were entitled, by such a principle, to assume that everything it was reasonable for him to believe at the outset would continue to be reasonable for him to believe throughout t_1–t_n, then given that it is reasonable at the outset to believe that there will be a surprise test at one of t_1–t_n, this would be reasonable to believe at t_{n-1} as well. Looking forward, S could once again exclude t_n on the grounds of the retained reasonable belief that there will be a surprise test at one of t_1–t_n. According to Wright and Sudbury, the temporal retention principle is the nub of the paradox, and once it is seen to be mistaken, the paradox disappears.[10]

However, I agree with Roy A. Sorensen that removing the temporal retention principle cannot as such be *the* solution, since a parallel case can be set up which eliminates the temporal dimension altogether. This is Sorensen's case of the Designated Student.[11] Five students are asked to stand in single file, each seeing the backs of those in front. A star is placed on the back of each by a teacher who announces, 'One and only one of you has a gold star on his back—the others are silver. But until you move from your present positions, the designated student will not be able to know (or be entitled to a reasonable belief) that he is the designated student.' The paradoxical reasoning can be reproduced by the student at the front (let's number them S_1–S_5 from front to back): 'Suppose the gold star were on S_5: he can see all the rest of us, so if he saw four silver stars he would know by a very simple deduction that he is the designated student. But the designated student cannot know he is the designated student until we all move out of line. Therefore, S_5 cannot be the designated student. Therefore, one of S_1–S_4 is the designated student. S_4 is as well placed as I am to do the reasoning I have just done. So S_4 knows that he is really the last student in the line who can be the designated student. But suppose the gold star is on S_4: he can see all the rest of us in who are front of him . . . etc.' This is a genuine version of the surprise text paradox. But the temporal retention principle is not required to generate the Thesis argument. (It requires an analogue, which is that whichever position in the line one is in, and whatever one sees, one has equal reason to believe the original announcement. This may also be a questionable principle. I return later to assumptions about the announcement's reliability for the participants in all these examples.)

[10] Ibid., p. 53.

[11] For the example, see Roy A. Sorensen, 'Recalcitrant Variations of the Prediction Paradox', *Australasian Journal of Philosophy*, 1982, pp. 355–62. (I do not discuss Sorensen's other two variations, but my proposed solution is, I believe, general enough to be able to deal with them.)

Sorensen's own proposed solution, in a later paper,[12] is also one that relies on the possibility of a surprise test on the last occasion. He develops the notion of an 'epistemic blindspot', which he explains as follows: 'A proposition p is an *epistemic blindspot* for a person a (at time t) if and only if p is consistent, while Kap (for a knows that p) is inconsistent.'[13] A 'conditional blindspot' he then goes on to define as follows: 'A proposition is a *conditional blindspot* for a (at t) if and only if it is not a blindspot but is equivalent to a conditional whose consequent is an epistemic blindspot.'[14] Sorensen then applies this to the surprise test: 'For the sake of simplicity, consider the 2 day case of the teacher's announcement involving only one student, Dave.'[15] The announcement is to be: 'Either the test will be given Thursday or Friday but in neither case will Dave know in advance.' Using my terminology, with 'E' standing for 'the test occurs', 't_4' for 'Thursday', 't_5' for 'Friday', and 'Kdp' for 'Dave knows that p', the announcement is equivalent to

(P1) $\neg (E \text{ at } t_4 \;\&\; \neg Kd\,(E \text{ at } t_4)) \rightarrow (E \text{ at } t_5 \;\&\; \neg Kd\,(E \text{ at } t_5))$

In other words, if the test fails to occur and surprise Dave on Thursday, it will occur and surprise him on Friday. Sorensen's claim is that while Dave can know this whole proposition—this is what he is genuinely apprised of by the announcement—nevertheless once he knows the antecedent (that a surprise test did not occur on Thursday), he is debarred from knowing the whole proposition. He cannot know 'E will occur at t_5 and I do not know it will'. In Sorensen's term, this is an epistemic blindspot for Dave once t_4 has elapsed. And the whole proposition (P1) is a conditional blindspot for Dave—though not for the onlooker who is not, in the original announcement, made a candidate for being surprised. The point is that someone in Dave's position, once he knows that it is not the case that the test occurs on Thursday, cannot infer that a test will occur and surprise him on Friday. And because he cannot infer that, once again, so the argument goes, it is possible for the teacher to surprise him with a test on Friday.

This account has one serious shortcoming: (P1) is not the right announcement for the teacher to make. By this I mean, firstly, that it is not the announcement made by the teacher in the original example that has so often been discussed. The announcement is simply that E will occur at one of t_1–t_n, and will be a surprise to S *when it occurs*. Sorensen substitutes for this an announcement that E will occur at one of t_1–t_n, and will be a surprise *at whichever of* t_1–t_n it occurs.[16] This is not the right announcement for the teacher to make for the more important reason that it fails to generate even the appearance of the required paradox. The teacher is supposed to commit herself from the outset to the possibility of surprising

[12] Sorensen (1984), referred to in n. 7 above. [13] Ibid., p. 131. [14] Ibid., p. 132.
[15] Ibid., p. 133. [16] This is similar to Ayer's substitution discussed in sect. II above.

S with a test at t_n and inform S of this. How, then, is S supposed to reason at the outset? It will now have to be on the basis of the following premisses: (*a*) there will be a test at one of t_1–t_n; (*b*) the test will be a surprise; (*c*) it is possible that the test will occur at t_n and be a surprise. But surely there is now not the slightest temptation for the student to construct the troublesome Thesis argument. If the student has as much reason to believe (*c*) as the others (and all are by hypothesis to be believed solely on the teacher's testimony), then the first step of the Thesis ('Suppose the test was on Friday: then before Friday I'd know it was going to be on Friday') is not even statable. The paradox only comes into being once we grant S the right to a belief in the negation of (*c*). So, if the announcement were what Sorensen here has it be, the paradox would be not so much solved, as rendered unstatable. And if we revert to the normal announcement, without (*c*)—that E will occur at one of t_1–t_n and S will be surprised by E when it occurs—we find no 'blindspot' for S. Once t_{n-1} has passed with no test occurring, S can know (have good reason to believe) that either there will be a test at t_n which does not surprise him or there will be no test.

The final recent attempt to defuse the paradox that I shall consider questions this last point, and hopes thereby once again to show that a surprise test at t_n is possible. This is the proposal of Doris Olin.[17] Olin sees correctly that the announcement gives S reason to believe just (*a*) and (*b*) above, namely: (*a*) that E will occur at one of t_1–t_n, and (*b*) that E will be a surprise to S (in her version, that S will not be justified prior to E's occurrence in believing that E will occur when it does). Her position, however, is as follows: S has equal reason for believing (*a*) and (*b*). Thus, supposing t_{n-1} to have passed and E not have occurred, then given that S has equal reason to believe both (*a*) and (*b*), S also has reason to believe 'There will be a test at t_n' (since he is justified in believing (*a*)) and 'I am not now justified in believing that there will be a test at t_n' (since he is justified in believing (*b*)). But while 'p and S is not now justified in believing that p' may well be coherent, it is surely not possible for S to be justified in believing it. 'Hence', Olin concludes, 'the student cannot justifiably believe the teacher's announcement on Thursday night and the argument is blocked at the very first step.'[18]

According to Olin, a consequence of this position is that even if the possible times are reduced to *one*, the prediction that S will be surprised can still come true. If the announcement is: 'E will occur at t, and you will not justifiably believe before t that E will occur at t', then Olin's analysis has it that S is justified both in believing that E will occur at t, and in believing that he is not justified in believing that E will occur at t. But since that is something that S cannot coherently believe, S is not justified

[17] Doris Olin, 'The Prediction Paradox Resolved', *Philosophical Studies*, 1983, pp. 225–33, and 'The Prediction Paradox: Resolving Recalcitrant Variations', *Australasian Journal of Philosophy*, 1986, esp. pp. 181–4.

[18] Olin (1986), p. 182.

in believing the announcement at all. Thus, when E does occur at t, S had no prior justification for believing that it would occur—that is, is surprised—and the announcement is thereby true.[19] There is something puzzling in this extreme variation. It is that the teacher has to announce something effectively of the form 'p but you do not now justifiably believe that p'. The notion that S has been given an authoritative testimony seems, consequently, to have collapsed. Or rather, S has it on testimony that the testimony itself cannot be relied upon. This, I think, is unfortunate, since Olin's strategy revolves around an assumption that S should be able to trust the teacher's testimony. It is at least unclear that this 'starkest form' of the surprise test paradox is a form of it all—unless, with Sorensen, we construe the announcement, for any value of n, as: 'E will occur at just one of t_1–t_n and (if E occurs at t_1, E will surprise you) and . . . (if E occurs at t_n, E will surprise you).' But I have argued that there are good reasons against doing this.

As I see it, however, the main problem with Olin's analysis is its explicit restriction to cases where S must rely solely on the announcement for all justification of any of his beliefs concerning E. This is supposed to ensure that S has equal justification for his belief that E will occur at one of t_1–t_n and for his belief that E will be a surprise. Without that assurance, the incoherent 'p and I am not justified in believing that p' cannot be generated. But since, as we have seen, the same paradoxical argument can arise in cases where S's belief in the two components of the announcement is far from equally justified, including some where S can know, independently of the announcement, that E will occur, the solution offered by Olin seems insufficiently all-embracing.[20]

IV

All the approaches I have considered in the previous section rely (despite their difference from Quine in other respects) on undermining the very first premiss of the Thesis argument, the premiss that if E is going to occur at t_n, then S will be able to know (justifiably believe) that E is going to occur at t_n. They all wish to say that prior to t_{n-1}, S has every reason to regard the announcement as entirely trustworthy, but that when only t_n remains, it becomes entirely untrustworthy, or at any rate wholly inaccessible as something S could justifiably believe. S's mistake, in constructing the Thesis argument, was to overlook this potential alteration of the announcement's status in one of the possible outcomes. But why is this not tantamount to saying that S's mistake was to regard the announcement as totally authoritative in the first place? If S should have foreseen that possibly he will have no reason to believe the announcement

[19] Olin (1983), pp. 230–1.
[20] This criticism is made by Chihara (1985), pp. 195–6.

at some time during t_1–t_n, then why insist that at the outset S must be allowed to assume that the announcement is totally authoritative? I shall hope to show in the remainder of this paper that a much stronger and simpler solution is that S cannot coherently assume the announcement to be totally authoritative. That sounds superficially like a return to Quine's proposal, but what I propose is not the same. For my proposal, unlike Quine's, works for versions of the paradox where S knows (or justifiably believes) *that E will occur*. It covers cases where S *can* regard the announcement as *authoritative in its claim that E will occur at one of* t_1–t_n, and cases where knowledge that E will occur is not conveyed by the announcement, but is independently available. In all these cases, S must be allowed the first premiss of the Thesis argument—that if E occurs at t_n, it will be no surprise. The solution does not, I suggest, depend on blocking that premiss. But nor does it depend on allowing S to regard the announcement as totally unauthoritative. Rather, my claim is that in all the variants there is one common assumption that S may never justifiably make: namely, the assumption *that E will be a surprise*. From S's point of view, it must always remain a possibility that E will not be a surprise to S.

To try to make some ground, let us turn to some simpler examples. Let us simplify matters enormously by removing any agency from the process leading up to the candidate surprise event, and by reducing the number of known times at which the event may occur to two. Suppose we have a hen that lays eggs, and that each egg it lays has an equal chance of falling down one of two tubes and thereby into one of two totally opaque, enclosed boxes B_1 and B_2. There is a single bell which rings whenever an egg falls into either box. I may open the boxes only in sequence. Let us call the time at which I open B_1, t_1 and the time at which I open B_2, t_2. Now, suppose that the bell rings, indicating that an egg has fallen. I know, prior to both t_1 and t_2, that there will be an event E of my seeing the egg. I know further that E will be at either t_1 or t_2.

By 'E will be a surprise to me' we must understand as before precisely the following:

(P) E will occur at some time, such that *prior to* that time I *do not know* that E will occur *at* that time.

In the 'egg' case, I know that possibly P. How do I know this? I know that E will occur at either t_1 or t_2, but, obviously in this case, prior to t_1 I do not know at which of t_1 and t_2 E will occur. So, prior to t_1, I know that possibly E will occur at t_1, *and* that if it does E will be a surprise to me. Hence I know that possibly P (or: possibly E will be a surprise to me).

This example, however, has another feature—namely, that if every possible time at which I know E may occur has elapsed, *bar one* (t_2 in this case)—then I can no longer be surprised, and I know then that I can no longer be surprised. After not seeing the egg in the first box, I know that it is in the second; and thereby I know that my seeing of the egg will not occur at a time such that prior to that time I do not know it will occur at that time. So initially, as well as knowing that E will occur at one of t_1 and t_2, and that possibly E will be a surprise to me, I know that if E occurs at t_2, E will not be a surprise to me. I know, therefore, that possibly E will *not* be a surprise to me. This is unproblematic.

Let us alter things now in a crucial respect. The bell has rung, and I have not yet looked in either box, but I am talking to someone who says, 'I've been collecting eggs for 35 years, and, strange as it may seem, whenever one finds an egg, it's a surprise to find it. You may safely assume that you will be surprised by your seeing of this egg.' Suppose that I am credulous and come to accept the following as true: 'I know that whenever E occurs, it will be a surprise to me', ('I know that P'). If I then hold fixed the premiss that I know that P, I might, prior to t_1, reason as follows:

1. I know that E will occur at just one of t_1 or t_2.
2. I know that whenever E occurs, E will be a surprise to me.
3. I know that if E does not occur at t_1, then prior to t_2 I will know that E will occur at t_2.
4. Hence I know that if E occurs at t_2, E will not be a surprise to me.
5. Hence, by 4, 1, and 2, I know now that E will occur at t_1.
6. But now it is prior to t_1.
7. So I know prior to t_1 that E will occur at t_1.
8. So I know that if E occurs at t_1, E will not be a surprise to me.
9. So, at whichever of the possible times E occurs, it is not possible that E will be a surprise to me.

Clearly this is an incoherent piece of reasoning. For 2 (I know that P) is inconsistent with the conclusion (it is not possible that P). So I must have been wrong to believe that I knew that the poultry expert's prediction would turn out true. Given 1 (I know that E will occur at one of t_1 and t_2), I can know that possibly P; I can also know that, if E does not occur at t_1, then E will not be a surprise to me. What I cannot know, or coherently believe myself to know, is that whenever E occurs, E will be a surprise to me. There is thus no antinomy here. The would-be Antithesis: it is possible that I will be surprised by E, is plainly true. But the would-be Thesis depends on inconsistent premisses, one of which I had no good grounds for accepting at all.[21]

V

Now we must reintroduce an agent of the same type as in the original surprise test case. This agent, T, not only predicts that E's occurrence will

[21] If we substitute 'have good reason to believe' for 'know' throughout the argument, the result commits the reasoner to the conjunction of 'I have good reason to believe that P' and 'it is not possible that P'. To regard this, in the context given, as a coherent argument for 'it is not possible that P' would be just as odd as regarding the original version using 'know' in this way.

be a surprise to me, but is fully in control of the timing of E's occurrence. Both these facts about T are known to me. However, so as not to move too fast, let us invent a second simpler example, which has in common with the 'egg' case the feature that there are only two times, t_1 and t_2, at which E can occur. Here is such an example.

T says to me (let us assume truly): 'I have a stone in one of my hands'. Let us allow further that by some means or other (which the reader may invent) I am enabled to know that a stone is in one of T's hands. T then also says the following: 'I will open my right hand at t_1 and my left hand at t_2. You will see the stone when I open the hand it is in. But when you see the stone, your seeing it then will be a surprise to you.' Is it possible for T's last prediction to turn out true?

Let us apply to this example what we have learned from the 'egg' example. E is now my seeing the stone in one of T's hands. I know in this case that E will occur at one of t_1 and t_2. So let us ask the following question: can the step 'I know that whenever E occurs, E will be a surprise to me' coherently occur in my reasoning when confronted with T? The answer to this question must be no. For, if this step is included, we can simply duplicate the incoherent reasoning of 1–9 above. In brief, if I allow myself to think: 'I know that whenever E occurs, E will be a surprise to me', then I can conclude, prior to t_1, that I know that E will not occur at t_2, hence that E will occur at t_1. Then, since I now know, prior to t_1, that E will occur at t_1, the prediction that E will be a surprise to me cannot be true. Once again, by including 'I know that P' in my reasonings, we arrive at the inconsistency: (I know that P) & (not possibly P).

Therefore, confronted with T's two hands, and the other assumptions we have introduced, I cannot reason on the assumption that I know A's prediction will come true. And without that assumption, nothing prevents the prediction's coming true. For, prior to t_1, I now have no way of knowing that E will not occur at t_2. And if I have no way of knowing that, I have no way of knowing that E will occur at t_1. So, if T opens her right hand at t_1 to reveal the stone, I will have been surprised by the occurrence of E. It is possible, therefore, that T's prediction will come true, that is, it is possible that I will be surprised to discover the stone in T's right hand.

The hypothesis stated above seems thus far to have been confirmed: given the other assumptions I have in these cases, I cannot make the assumption that I know I will be surprised by E; and if I cannot make the assumption that I know I will be surprised by E, it is possible that I will be surprised by E. For we have as yet seen nothing else that will allow me to know that E will not occur at the last of the times in a given series.

VI

Surely things cannot be so simple, it may be objected, even in the 'stone' example. In the 'stone' example, we relied on this point: that, prior to t_1, I cannot know that E will not occur at t_2. This was crucial, since if I can know from the start that E will not occur at t_2, I can know from the start that it will occur at t_1, and then the game is up for T's prediction that I will be surprised. But we have forgotten to mention recently something that was mentioned earlier in the 'egg' example, namely this: from the very beginning, prior to t_1, I know that if E has not occurred at t_1, then E will occur at t_2. Knowing this enables me to reason that if E does not occur at t_1, then E cannot surprise me. This was an irrelevant consequence in the 'egg' example, but surely it cannot be irrelevant now. For if I am entitled to assume that T is rational, is trying to surprise me with the occurrence of E, and has truly told me that E will occur at either t_1 or t_2, then am I not entitled to infer that T will not make E occur at t_2? Any rational person would know that, as things are set up, the prediction that E will surprise me cannot come true if E is left to occur at t_2. So I can reason that T would know that E's occurring at t_2 would not surprise me, and so would not leave the occurrence of E until t_2.

The problem now is this: if we allow the reasonable assumptions that T is rational and sincere, the argument for the Thesis of the antinomy reappears. If I assume T to know that I will not be surprised if E occurs at t_2, and sincerely to intend that I be surprised, then can I not know that E will not occur at t_2? If I can, I can know that E will occur at t_1, in which case there is no possible time at which E's occurrence may surprise me.

However, think back to the 'egg' example. There, uncritical acceptance of the assumptions that my informant was sincere and rational led me into incoherence. Crudely, the argument used there for the Thesis that I could not be surprised by E was of the form 'I know that P, I know that Q, ..., therefore not possibly P'. If, in the 'stone' example, the reasoning has to be of the same form in order to get the Thesis off the ground, then the correct conclusion must be that I am not entitled to assume T's full sincerity and rationality. If that assumption leads me to reason on the basis of the assumption that I know E will not occur at t_2, then, given my other assumptions, I will reason incoherently.

Thus, in the stone example, to be coherent I must allow the *possibility* that T is either bluffing or has made a startling error of reasoning. Either case leaves open the possibility that E will be in the second hand and hence that I shall not be surprised. And it is precisely because I must leave open this possibility, that it is possible for T, sincerely and rationally, to predict my surprise, and to be correct in her prediction. The Antithesis is true, the Thesis false; there is no antinomy.

Of course, it would not be very satisfying to take part in this trick with the stone, for either party. There is only one possibility of my being surprised, and it is highly unlikely that T will overlook this fact. E is considerably less likely to occur at t_2 than at t_1. Though 'technically' E's occurrence at t_1 surprises me—I did not know prior to t_1 that it would occur at t_1—it is, psychologically speaking, not very surprising.

But now if we increase the series of possible times again, and return to the case of the coin in one of five boxes, as described in section II, we can see how genuine surprise is possible, and can be predicted by the person who sets up the order in which I open the boxes. For we have a series of five possible times, t_1–t_5, at which E may occur. I still know—as Quine thought I should not—that, if E has not occurred at one of t_1–t_4, then it will occur at t_5. But this merely makes it vastly more probable that E should occur at one of t_1–t_4 than at t_5. And because of this, if E has not occurred before t_4, I will know before t_4 that t_4 is the only remaining possible time at which I may be surprised. So, it is again more probable that E should occur at one of t_1–t_3 than that it should occur at t_4.

Assuming that this argument is repeatable back through the whole series, then, as I survey the five unopened boxes, the probability of the coin's presence would seem to diminish along the line from B_1 to B_5. But it does not diminish to 0: the series of what I must regard as possible times of occurrence for E continues to have five members.[22] Given this, even if I have some way of calculating the relative probabilities of the coin's being in B_1, B_2, or B_3, it seems to have become a betting matter which box it is in. It seems that, if the series t_1–t_n is long enough, I can be not only 'technically', but 'genuinely' surprised by E's occurrence. Any account which has this consequence is surely to that extent attractive. For a very basic intuition about the possibility of achieving a surprise which has been announced in advance is that the greater the number of possibilities of the event's occurring according to the announcement, the clearer the possibility of a surprise. If we take as our example the two day case: 'Either today or tomorrow there will be a test, and it will be a surprise to you', then there is scarcely a prior intuition that *obviously* it must be possible for the announcement to be true. At the other extreme, if we imagine a teacher who says, 'On one of the next two hundred days there will be a test, and it will be a surprise to you', we will intuitively regard with enormous suspicion anything which suggests that this announcement could not be fulfilled. It seems overwhelmingly obvious that it can be fulfilled, and the fact that one can know from the outset that the two-hundredth day has almost no chance of being picked is totally uninteresting. The obviousness of the possibility of a surprise in line with the announcement is thus a matter of degree. My account handles this feature succinctly.

For the reasons given, then, the prediction that I will be surprised can

[22] Cf. Gardner (1963), p. 152.

come true, and there remains no argument at all for thinking that it cannot.[23] What is perhaps surprising about the case is that a simple prediction, which is rationally and sincerely made, can be seen to be possibly true if and only if it is regarded as possibly false.

VII

Many recent writers, including those discussed above in section III, have attempted to fix the conditions of the surprise test example by stipulation in such a way that my solution is blocked from the beginning. My reply to this is that we must recognize that there are limits to what we can coherently stipulate about such examples. And a final strength of my proposal, I suggest, is that acknowledging this limitation is the only price we have to pay for solving the paradox.

One stipulation that has been made is that the announcement which the teacher makes must be understood as genuinely informative. As Wright and Sudbury put it, any account which aims to solve the paradox '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.'[24] It might be thought that my proposal suffers by transgressing this requirement. For I have in effect said that the pupils cannot legitimately take themselves to have been informed, by the announcement, that the test will surprise them. However, I can allow that they have been informed that there will be a test on one of the stated occasions, and informed of the teacher's expressed intention to surprise them. This, together with background assumptions on their part about the teacher's likely sincerity and likely competence, is sufficient for them to believe that a surprise test is highly likely on one of the stated days. I suggest that this is sufficient, in a very ordinary sense, for them to be informed by the announcement. Certainly 'getting on with one's revision' is the appropriate action if this is what one is apprised of. Even a student who 'notices no peculiarity' is implicitly informed, if you like, that it is highly unlikely that the test will occur on the last day—all concerned can be allowed justifiably to believe this, or even to know it, without the threat of paradox. So I maintain that I can both solve the paradox and meet the requirement that the teacher's announcement be informative.

But is the talk of 'likelihood' a subterfuge here? Surely, someone may object, I have in effect admitted that the pupils, in being informed by the announcement, will be justified in believing that they will be surprised. All

[23] I have discovered that the essential feature of the solution I propose—the inadmissability of *S*'s claim to know that the prediction of a surprise will be true—is succinctly stated by T. H. O'Beirne, *New Scientist*, 25 May 1961, pp. 464–5. I owe the reference to Gardner (1963), p. 152.

[24] Wright and Sudbury (1977), p. 42.

we have to do now is to stipulate that all the pupil's beliefs about the test are *as* justified as this belief is, and my solution is demolished. It may work for those cases considered earlier, in which S's justification for believing (A) that E will occur at one of $t_1 - t_n$ is greater than his justification for believing (B) that E will be a surprise; but if the justifications are stipulated to be equal for both beliefs, my analysis fails in the central case. This stipulation of equality of justification is made by Olin,[25] and indeed figures as a crucial premiss in her proposed solution:

Would the student be justified in believing (A) on Thursday evening? I think not. For suppose he were so justified. Now surely he would be justified in believing (A) only if he were also justified believing (B), for there is no epistemically relevant difference for him between the two propositions.[26]

And, as Olin goes on to say, if he were equally justified in believing both (A) and (B), he would be justified in believing something of the form 'p and I am not now justified in believing that p'—which, as we have seen, leads her to say that the student could not at that stage be justified in believing *either* (A) or (B).

It certainly looks as if my solution works for cases where the justification for believing (A) somehow exceeds the justification for believing (B), whereas once the justifications are stipulated to be equal, Olin's analysis works, and mine does not. Notice that there are no cases of the surprise test paradox where S's justification for believing (B) is greater than that for believing (A). If one's grounds for believing that there was going to be a surprise test were more secure than one's grounds for believing that there was going to be a test at one of $t_1 - t_n$, then one could not generate the supposed paradoxical argument of what I have called the Thesis. The surprise test *might*, after all, be in the following week. If, on the other hand, one's beliefs are (A) that there will be a test at one of $t_1 - t_n$, and (B*) that there will be a surprise test *at one of* $t_1 - t_{n-1}$, then clearly belief in (B*) is parasitical on belief in (A), in a way which suggests that I could not be more justified in believing (B*) than in believing (A). (A parallel would be the pair of beliefs 'There is someone in the next room whom I have never met' and 'There is someone in the next room'. I could not be more justified in believing the former than in believing the latter. Reasons for believing the former must include, and cannot be exhausted by, reasons for believing the latter.) There are, as we have seen, plenty of cases where justification for belief in (A) can exceed justification for belief in (B), and my solution deals with them satisfactorily. There are no cases, I have maintained, where the reverse is true—where belief in (B) is more justified. The only apparent problem for me, then, is with the cases of equal justification.

[25] Cf. Olin (1983), p. 228, Olin (1986), p. 182.
[26] Olin (1983), p. 228.

What I think we must question is whether the stipulation of equal justification is a natural, and ultimately a coherent one to make. First a general query: is it, generally speaking, a natural assumption that if someone informs me in a single utterance that p and that q, then I have equal justification for believing that p and that q? Olin says that there is no epistemically relevant difference between (A) and (B) for the student for the following reason: 'Since both are based just on the teacher's testimony, his evidence for one is as good as his evidence for the other.'[27] But even a reliable informant need not be *equally* reliable across the board. Consider the following announcement: 'It is raining now and it was raining here precisely three hundred years ago.' My informant is reliable, and informs me of two propositions. We may suppose, if we like, that her utterance is totally true, and that I have no other evidence for believing in either component. Do I *thereby* come to have an equal justification for believing both components of her utterance? It is hard to see how I do just like that, given the different nature, on fairly ordinary assumptions, of *my informant's* justifications for the two beliefs she expresses. And in the case of the announcement in the surprise test case, such an assumption—that because the single source of both beliefs (in (A) and in (B)) is one announcement, or even one announcer, they are attended by an equal justification—is similarly dicey. For with the announcement that E will surprise me, I have to take into account that I am required (*a*) to believe something about what will happen in the future, (*b*) to believe something, on someone else's testimony, about one of my own mental states, (*c*) to believe on someone's testimony that something that they intend to bring about will come about. All these are factors which can affect the degree to which I may regard a testimony as reliable. It would seem odd to assume as a general rule that if someone informs of me of (B) (that E will surprise me) and of any other proposition p, then there is 'no epistemically relevant difference' for me between p and (B). And if p is a proposition lacking any one of the complicating factors (*a*)–(*c*), we surely have prima facie grounds for doubting whether there is truly 'no epistemically relevant difference' between p and (B).

It begins to look, then, as if a stipulation merely that (A) and (B) are both believed by S solely on T's testimony will not guarantee that S has equal justification for believing both (A) and (B). What more can we add by way of a stipulation which will entitle S to assume himself equally justified in believing (A) on (B) on T's testimony? It is common to require that the participants in the surprise test example be reliable informants and perfect logicians, or 'ideal thinkers in favourable epistemological environments'.[28] Let us accept this. We can even go on to spell out that the teacher is a perfect logician incapable of any deceit and in possession of

[27] Olin (1986), p. 182. [28] Sorensen (1984), p. 131.

every relevant fact, so that whatever she says is equally—because fully—reliable. I shall not quarrel with any of these stipulations if anyone wishes to make them. But where we must draw the line is at stipulating that the student should know or be justified in believing all of this about the teacher. Or rather, more precisely, I draw the line at stipulating *both* that the student know or be justified in believing all this about the teacher *and* that the student himself be an ideal thinker. If we do not draw this line, we are attempting to describe an impossible situation.

To see this, let us try to reconstruct the paradox with all the required assumptions made fully explicit. We assume that T is an ideal thinker and uniformly tells the truth, that T makes the usual announcement about a surprise test, that S is an ideal thinker, and that S justifiably believes that T is an ideal thinker and uniform truth-teller. Now we have S reason roughly as follows:

(1) T says that (A) E will occur at one of t_1–t_n, and that (B) E will be a surprise to me.
(2) T is a uniform truth-teller.
(3) Therefore, (A) and (B) are true.
(4) T is an ideal thinker.
(5) Therefore, T knows that if E occurs at t_n, then (B) is not true.
(6) Therefore, E will not occur at t_n.

etc.

Ultimately, as before, S will conclude by such reasoning that (B) is not true—E will not surprise S. But we have assumed that S is an ideal thinker, so let us continue to insist on this stipulation. In consequence, S must be allowed to draw this further conclusion from his reasoning:

(T says that (A)) and (T says that (B)) and (T is an ideal thinker) and (T is a uniform truth-teller) and not-(B)

No ideal thinker could fail to view this as a consequence of his prior reasonings, but no ideal thinker could regard this as an argument for the proposition, not-(B). No moderately good one could do so. So, if the paradox is to be statable, either S must be far less than an ideal thinker or S must be robbed of at least one of the assumptions we stipulated him to use. Of course he must continue to believe that T announces that (A) and that (B). But, I suggest, *one of* 'T is an ideal thinker' and 'T is a uniform truth-teller' must be absent from S's repertoire of justified beliefs, if S is an ideal thinker.

We thus come up against the limits of what can be stipulated.[29] There

[29] Chihara (1985) makes a comment in similar vein when he says that the student in the imagined situation 'cannot say "I know the initial premiss is true because I know that you have stipulated this to be the situation." That would be absurd!' (p. 197).

simply cannot be two ideal thinkers, one of whom announces the standard surprise test to the other, where the other justifiably believes the first to be a uniformly truth-telling ideal thinker. Whether this goes against any strong intuitions, I cannot say. If it does, I submit that learning this somewhat recherché truth and adjusting our intuitions accordingly is a small price to pay for resolving the supposed paradox. The consequence is that, in the case where there is no justification for belief in (A) independently of the announcement, the most *S* can be justified in believing is *either* '*T* is an ideal thinker and probably is not bluffing in saying that (A) or in saying that (B)' *or* '*T* is absolutely incapable of deceit and probably has not made a logical error in saying that (A) and (B)'. Armed with either of these beliefs, or with the disjunction of them both, *S* can never discount the last occasion provided for by the announcement as a possible time for the test to occur.

Finally, if we now choose to stipulate that *S* believe (A) and (B) with equal justification, this can only mean that, given *T*'s testimony, *S* must think it *as likely* that (A) is true as that (B) is true. But here, as with the cases of unequal justification, it is amongst other things the case that, from *S*'s point of view, (B) might be false. In addition, in the case of stipulated equal justification, (A) may also be false from *S*'s point of view. But it is more important that in the case of stipulated equal justification, and in all the other cases considered, *S* must regard (B) as possibly false. This gives us a single straightforward solution for all the cases of the paradox.[30]

Department of Philosophy CHRISTOPHER JANAWAY
Birkbeck College
Malet Street
London WC1E 7HX

[30] I am grateful for the help of Dorothy Edgington, Simon Blackburn, and Sam Guttenplan.