RICHARD L. KIRKHAM

THE TWO PARADOXES OF THE UNEXPECTED EXAMINATION

(Received 22 April, 1985)

What makes the paradox of the unexpected examination more than just a philosophical curio is the fact that, if it cannot be dissolved, it entails that there is something drastically wrong with standard logic. (Some students of the paradox have reached just this conclusion.)[1] Thus a dissolution of the paradox is a proof that the paradox does not constitute a counter-example to the principles of standard logic. As a practical matter, this means filling in the missing steps in the student's argument (see below) and showing why it is unsound. I shall argue that at least one reason why scholars of the paradox cannot agree on the correct dissolution is that there is no *one* correct dissolution. The brief story of the unexpected examination hides within it at least two different paradoxes, each with its own solution. We can reveal one or the other by the way that we fill in the details of the story. And it is to the story that I now turn:

One day a teacher tells his class

> I am going to give you a surprise examination sometime next week, very possibly on the last day. As you know I do not believe in announcing the date of examinations in advance. For this reason I shall never give you an examination which does not come as a surprise. Accordingly, I am not going to tell you which day next week I shall be administering the test, except that it will be on one of the five weekdays. You will not find out until I actually hand out your examination questions that the test is scheduled for that day. You will be surprised, even if I administer it on the last weekday.

At that point a bright student in the back of the class says:

> But, sir, you cannot possibly give us an examination next week: If you have not given us an examination by Thursday, then we shall expect on Thursday night that the examination will be the next day; hence, we will not be surprised when you administer it. But you have said that you will never give us an examination that does not come as a surprise, so you cannot give us the examination at all on Friday. On the other hand, you cannot give us an examination on Thursday; for, believing that it cannot be on Friday, we would believe by Wednesday night that it must be given the following morning, so, if it comes, it will not be a surprise. For similar reasons you cannot give us a surprise examination on Wednesday, Tuesday, or Monday. Therefore, since you will not give us an un-surprising examination, you cannot give us any examination at all.

Philosophical Studies **49** (1986) 19–26.

The first thing the following Friday morning, the teacher gave his students an examination and they were surprised.

There are many variations of the basic story. The "know" variation has the bright student's second sentence read "...then we shall *know* on Thursday night...etc." The "deducibility" variation has the same clause read "...then we shall be able to deduce on Thursday night...etc." The weakness of these renditions as compared to my own "expectations" version is that they attribute to the student premises about what he knows or can deduce that are highly dubious. So much so that these renditions come dangerously close to saying that the student made an *obvious* mistake and, hence, there is no paradox here in the first place. But there is. We *are* "taken in" by his argument. Whatever he does wrong it is surely not an *obvious* mistake. For example, one school of thought on the paradox, represented by Doris Olin and Charles Chihara among others, holds that the story reveals an incompatibility between logic and certain epistemic theorems. Chihara says that the student's argument is "... a *reductio ad absurdum* of the assumption that the students know that what was announced is true. This is a paradoxical result since we seem to have very strong intuitions that the students could very well know just that."[2] But *I* have very strong intuitions that the students *could not know* (as distinct from merely *expect*) that the announcement is true.

How can it be that an apparently sound argument shows that the announcement cannot come true when in fact, as we shall see, it can? It seems that reality has slapped logic in the face. But as it stands, the student's argument is ellyptical. Brian Medlin fills in the argument in such a way that the premises of the student's argument are seen to be viciously self-referential.[3] (Vicious, here, means that a contradiction can be derived therefrom.) So his dissolution of the paradox is to point out that the story is not really a counter-example to logic: For this is *not* a case in which a genuinely sound argument has generated a false conclusion.

But suppose there is a second way of filling out the student's argument which contains no self-reference? To rescue logic we would need to find some way in which this second argument were unsound. I do both below. However, this does not mean that Medlin's dissolution has failed; for we not only have two dissolutions here, we have two different paradoxes. I say this because the flaw that produces unsoundness in my reconstruction of the argument is not present in Medlin's version, just as his problem of self-reference is not present in mine. Were that not the case, I would be in a position to say that I had identified a flaw in *both* versions and he had not; hence, I had successfully protected logic and he had not. Since, as will become clearer below and from reading Medlin's article, neither reconstruction suffers from any other failures of soundness beyond those he and I point out, there is no flaw common to both. Thus, each reconstructed argument, were there not someone to point its flaw, would stand as a distinctly different *prima facie* counter-example to standard logic. Different because finding the flaw in one would not entail that there was a flaw in the other. Hereafter, "the paradox" will refer to this second paradox which does not turn on a problem of self-reference.

Here, then, is my reconstruction of the student's argument:

Key Premise (KP): We (the students) expect an exam on or before Friday of next week.

(1) There will be no exam before Monday.

(2) You (the teacher) will succeed in your intention *not* to give us an *unsurprising* exam.

(3) For the whole of next week we shall always have correct beliefs about what day of the week it is.

(4) For the whole of next week we shall always have correct beliefs about whether or not we have been given an exam so far in the week.

(5) If a person expects event E to take place on one of a finite set of times S; then: if he believes all but one of the times in S have passed and he believes that E has not taken place, then he will expect E to take place on the remaining time in S.

(6) If we expect an exam on or before Friday; then: if we believe it is Thursday night and we believe the exam has not been given, then we shall expect an exam on Friday. (instantiation of (5))

(7) Therefore, if we believe it is Thursday night and we believe that the exam has not been given, then we shall expect an exam on Friday. (from KP and (6) via *modus ponens*)

(8) If no exam will be given on or before Thursday, then (on Thursday night) we shall believe that it is Thursday and that the exam has not been given. (from (3) and (4))

(9) Therefore, if no exam will be given on or before Thursday, then (on Thursday night) we shall expect an exam on Friday. (from (7) and (8) via the Law of Hypothetical Syllogism)

(10) If an event E will take place at time T, and E is expected (in advance) to take place at T, then E will be an unsurprising event.

(11) If an exam will take place on Friday, and the exam is expected (on Thursday night) to take place on Friday, then the exam will be an unsurprising exam. (instantiation of (10))

(12) Therefore, if no exam will be given on or before Thursday, and an exam will be given on Friday, then there will be an unsurprising exam on Friday. ((9) and (11) via the Law of Substitution)

(13) There will *not* be an unsurprising exam on Friday. (from (2))

(14) Therefore, it is *not* the case that *both* no exam will be given on or before Thursday *and* an exam will be given on Friday. (from (12) and (13) via *modus tollens*)

(15) Therefore, if no exam will be given on or before Thursday, then no exam will be given on Friday. (from (14) via *modus tollendo ponens*)

From this point, (15) seems to yield:

(16) If there will be no exam on or before Thursday, then there will be no exam at all next week.

By a series of parallel steps for each day of the week, the bright student tries to arrive at:

> If there will be no exam on or before Sunday, then there will be no exam at all next week.

This last premise, plus (1) above, yields via *modus ponens* his ultimate conclusion: There will be no exam at all next week. Of course, it would be no easy trick to get from (16) to the next parallel step:

> If there will be no exam on or before *Wednesday*, then there will be no exam at all next week.

But, as I shall show below, any mistakes the student makes after (16) are not relevant to the dissolution of the paradox. Accordingly, I shall not fill out the remainder of the argument.

Note that the unargued premises are KP, (1)–(5), and (10) ((6) and (11) are intended to be fair instantiations of (5) and (10) respectively). There cannot be anything unsound in the student's assumption of (1)–(4) because every one of these comes true in the story, and, as I shall argue below, the story does describe a logically possible set of events. I take it that (5) and (10) are more than plausible. (I know of no one who would deny either.) Hence, the unsoundness must come from KP.

In particular note that none of the student's premises refers to itself or to any of the other premises. And none of the premises is self-contradictory nor can a contradiction be deduced from them in contrast to those who have attributed inconsistent premises to the student.[4] What usually leads people to think that the student has contradicted himself is the mistaken belief that he must assume "there will be an exam next week" when, in fact, all he need assume is KP. And while, as I shall show, the negation of KP is true when *and only when* the student finishes making his argument, it (the negation of KP) cannot be *derived* from the permises.

But, of course, there is *something* wrong with KP which reads "We expect there will be an examination next week." Intuitively, we can see that the premise is logically crucial to the bright student's argument, for without it, the mere fact that the students had not had an examination on the first four days of the week would give them no reason at all to expect an examination on Friday. Hence, the bright student's intermediate conclusion that any examination on Friday will not come as a surprise depends on the KP and, thus, so does the rest of his argument. Why is it easy for the bright student (and for us) to assume this premise? Because *the key premise is absolutely true at the time that the student begins his argument*. At that point the students *do* expect an examination next week, for the teacher has just told them that there will be one (and, thus, contrary to Quine,[5] the expectation is warranted). But, and here is the key to the whole paradox, after the bright student completed his argument, *the argument itself convinced the students that there would not be an examination the following week and, so, it changed their expectations in such a way as to render the hidden premise false*: The argument renders itself unsound by virtue of its own persuasiveness.

Statements whose truth value changes over time are familiar enough. Most contingent statements not indexed to a particular time are like this. (E.g. "my shirt is blue" is true today, but was false yesterday.) *Usually*, they can be used as premises without causing any problems. But for those who are made uncomfortable by this characteristic, note that it is not essential to either the student's argument or my dissolution. To freeze the truth value of KP I simply add the appropriate indexical clause: "(KP_1) In the time between your (the teacher's) announcement and by beginning my argument, we expect an exam on or before Friday of next week." Then I add to the argument a second key premise: "(KP_2) Our (the students') expectations will not change between now and Friday." KP_1 starts true and stays true. KP_2 starts false and stays false. So, the essence of dissolution is still here: The student based his argument on the expectations of his peers and these expectations change as a result of his argument. However, there is non-logical advantage for using KP instead of KP_1 and KP_2: Since KP_2 is false right from the start we are left with an unanswered question. Why does the student assume it and why are philosophers trained to spot this sort of thing so easily taken in by it? With KP we have an explanation: At the time it implicitly enters the argument it is true.

Let me now quickly discharge two promisary notes I issued earlier: First, it has been alleged that the story of the unexpected examination is a logically impossible scenario: Given the teacher's announcement, the students simply could not be surprised. (Which would mean he contradicted himself in calling the upcoming examination a "surprise".)[6] But it should now be clear that despite the teacher's announcement and their initial inclination to expect an examination, it is quite possible that the students can come to expect that there will not be an examination the next week and, thus, be surprised when it comes.

Secondly, it has been contended that the bright student's argument is perfectly sound through the intermediate conclusion that there can be no surprise exam on Friday. Only when he tries to extend this to the rest of the week does he go wrong.[7] In fact, whatever is wrong with the student's argument goes wrong in the very first part; for there *is* a surprise examination on Friday and this is a logically possible situation. Moreover, it is quite possible to construct a version of the paradox in which the teacher actually specifies on what day the surprise examination will take place. He tells his students that there will be a surprise examination on the following Monday. A student argues that this cannot happen since the students will, on Sunday night, expect an examination for the following morning. So, given the teacher's intention not to give an *un*surprising examination there will in fact be no examination Monday. The next Monday morning the students are surprised to get an examination. This, too, is a logically possible situation, for the argument would cause a change in the students' expectations. The fact that a paradox can be generated even when the teacher specifies the exact day of the examination is significant: Any resolution which claims that the student's argument goes wrong only *after* he has soundly deduced that there cannot be a surprise examination on Friday would have to allow that there is nothing at all wrong with the student's argument in this "exact day" version. But given that his conclusion is falsified, there *must* be something wrong with it. This point provides further evidence that the paradox here is different from the self-reference paradox dissolved by Medlin, for it is not obvious that there could be a logically possible "exact day" version of the self-reference paradox.[8]

A further implication of my resolution is that any theory of expectation which holds that expectations are never changed by deduction must be recognized as an idealization which may create paradoxes.

NOTES

[1] The belief that the paradox is undissolvable and, hence, a genuine counter-example to logic has not been popular of late, but while it was it produced some of the most exotic claims in the history of philosophy. It has been said that the paradox shows: that the Law of Non-contradiction is not always true; that "It is true that x or not-x" does not entail "It is true that x or it is true that not-x"; that contradictions can be true; that one can derive a contradiction from a statement which is not itself self-contradictory; and that some announcements about contingent future events can guarrantee that the event will take place. See Paul Weiss, 'The prediction paradox', *Mind* 61 (1952), 265–9; Martin Edman, 'The prediction paradox', *Theoria* 40 (1974), 166–77; G. C. Nerlich, 'Unexpected examinations and unprovable statements', *Mind* 70 (1961), 503–13; and Michael Scriven, 'Paradoxical announcements', *Mind* 60 (1951), 403–7.

[2] Charles S. Chihara, 'Olin, Quine, and the surprise examination', *Philosophical Studies* 47 (1985) 191–199, especially 191; and Doris Olin, 'The prediction paradox resolved', *Philosophical Studies* 44, 225–233.

[3] Brian Medlin, 'The unexpected examination', *American Philosophical Quarterly* 1 (1964), 1–7.

[4] See, for example, J. M. Chapman and R. J. Butler, 'On Quine's "So-called paradox"' *Mind* 74 (1965), 424–5; Igal Kvart, 'The paradox of the surprise examination', *Logique et Analyse* 21 (1978), 337–44; James McLelland and Charles Chihara, 'The surprise examination paradox', *Journal of Philosophical Logic* 4 (1975), 71–89; and Judith Shoenberg, 'A note on the logical fallacy in the paradox of the unexpected examination', *Mind* 75 (1966), 125–7.

[5] W. V. Quine, 'On a so-called paradox', *Mind* 62 (1953), 65–7.

[6] See, for example, L. Jonathan Cohen, 'Mr. O'Connor's "Pragmatic paradoxes"', *Mind* 59 (1950), 85–7; Edman, *op. cit.*; Ardon Lyon, 'The prediction paradox', *Mind*

68 (1959), 510–17; Nerlich, *op. cit.*; D. J. O'Connor, 'Pragmatic paradoxes', *Mind* 57 (1948), 358–9.

[7] See, for example, Edman, *op. cit.*; Lyon, *op. cit.*; R. A. Sharpe, 'The unexpected examination', *Mind* 74 (1965), 255; Shoenberg, *op. cit.*; and J. A. Wright, 'The surprise exam: Prediction on the last day uncertain', *Mind* 76 (1967), 115–7.

[8] Medlin, *op. cit.*, but especially R. Shaw, 'The paradox of the unexpected examination', *Mind* 67 (1958), 382–4.

Dept. of Philosophy,
University of Notre Dame,
Notre Dame, IN 46556,
U.S.A.