DORIS OLIN

THE PREDICTION PARADOX RESOLVED

(Received 10 October, 1982)

The prediction paradox, usually expressed in terms of a surprise examination, continues to baffle and charm. No generally accepted solution to this paradox has yet appeared. That this knot in our reasoning should resist treatment is disconcerting; it would, at the very least, be reassuring to have it untied. In this paper, I attempt to do so by offering a new solution – a solution which suggests that the paradox has philosophical significance in the realm of epistemology and is closely related to the well-known lottery paradox.

The paradox: A teacher announces to his student, S, that he is going to give him exactly one examination sometime during the next week, and that it will be a surprise – S will not be able to predict, prior to the day of the examination, on which day it will be held. The student claims that this is impossible. He argues as follows: 'If the examination were held on the last day of the week, then on the previous evening I would be able to predict that the examination would be given the next day. But if the examination were held on the fourth day, then, since an examination on the last day would not be a surprise, I would be able to determine on the previous evening that the examination would be given on the fourth day. Similarly for each of the remaining days. So the surprise examination cannot be given on any day of the week and it is therefore impossible for the announcement to be satisfied'.

The teacher, we would hope, is unconvinced. For it is clear that he can give a surprise examination on, say, the second day. Thus we have a paradox; a seemingly air-tight argument leads to a conclusion which is patently false.

What does resolution of the paradox require? In general terms, a paradox arises when there appear to be strong reasons to believe each of two inconsistent propositions. Certainly it does not suffice merely to adduce further arguments for one of the two propositions. For example, it is pointless to argue, in the present case, that since the student has a good argument to show that the teacher cannot keep his word, he cannot believe the teacher; consequently S has no reason to expect an examination during the next week;

Philosophical Studies **44** (1983) 225–233. 0031–8116/83/0442–0225$00.90

and, hence, an examination at any time during the week will be a surprise.[1] What is needed is not a proof that a surprise examination is possible; that is something we already know.

In general, to dissolve a paradox it is necessary to undermine or defuse the argument for the false proposition – to reveal the fallacy or error in the argument. In this case, since a surprise examination clearly is possible, we must diagnose and isolate the errors in the student's notoriously slippery argument.

Uncovering the flaw in the student's reasoning requires that we set out explicitly the implicit assumptions on which his argument depends. But we must take care to avoid formulating premises stronger than those the argument really needs; to rebut such premises may leave the paradox untouched. We must, accordingly, try to find the minimal, least problematic set of premises sufficient for the task. It is worth bearing in mind that the paradox generating argument is one we all find seductive; hence we are likely to be off the mark if our solution consists in isolating and refuting a premise that is implausible or highly controversial.

Perhaps the single most widely accepted approach to the problem is to construe it in purely logical terms, by interpreting the teacher's statement in terms of deducibility. The problem arises, it is suggested, because the student *mistakenly* construes the teacher's announcement as:

(T) There will be exactly one examination next week and it will be held on a day D such that *this proposition* and the fact that the examination has not been held on any day prior to D do not jointly entail that the examination will be held on D.

It can be shown, however, that (T) does in fact entail a contradiction. So the teacher cannot, if what he says is possibly true, mean (T). Typically, it is claimed that the correct interpretation is a revised version of (T), suitably modified to avoid contradiction.[2]

The difficulty with this general approach lies in the suggestion – mistaken, in my opinion – that the student's argument results from construing the teacher's announcement as (T). Notice that the teacher, as well as the student, can work through the paradox generating argument. But he can do so only if he tacitly assumes that the surprise examination has been previously announced to the student. Without this tacit premise, the argument cannot begin. The argument that (T) cannot be satisfied, however, needs no such assumption.

The announcing of the surprise examination is crucial to the problem because it must be plausible to suppose, at certain points in the argument, that the student would have good reason to believe that a surprise examination will be given. And the sole reason for thinking that the student would be entitled to believe this is that the teacher, who is generally reliable, has said so. Epistemic concepts such as these are, I believe, crucial to the paradox.

What exactly are we claiming when we say that the examination will be a surprise? The relevant issue, it seems, is what the student is *justified* or *warranted* in believing. That is, to say that the examination will be a surprise is to say that the student will not be justified in believing, before the day of the examination, that the examination will occur on that day. With this epistemic interpretation of the key concept, it is now possible to isolate the essential errors in the paradox generating argument.

First, we must be a little more specific about the details of the case. The teacher, let us suppose, said to S: 'An examination will be held on exactly one of the days Monday–Friday; and if an examination is in fact held in the afternoon of day D, you will not be justified in believing this before that day'. Now if the student's argument is not to be open to trivial objections, at least the following assumptions concerning S's memory, reasoning powers, etc. are needed. (P_1) The student is an expert logician; so that if he were justified in believing $p_1, ..., p_n$ which jointly imply (or strongly confirm) q, then he would see that $p_1, ..., p_n$ jointly imply (or strongly confirm) q. (P_2) On Sunday evening, and throughout the next week, the student remembers that the teacher is generally reliable and, also, remembers what he said. (P_3) On Sunday evening, and on any evening of the week, the student knows what evening it is, and on any evening of the week, he remembers whether an examination has been held on that or any previous afternoon of the week.

As well as these assumptions concerning the details of the particular situation, the argument must also rely on certain general epistemological premises. Roughly, these may be stated as follows. (P_4) If A is justified in believing $p_1, ..., p_n$, $p_1, ..., p_n$ jointly imply q and A sees this, then A is justified in believing q. (P_5) If A is justified in believing $p_1, ..., p_n$, $p_1, ..., p_n$ strongly confirm q, A sees this and has no other evidence relevant to q, then A is justified in believing q.

Now even assuming the details concerning *S*'s memory and reasoning ability, it is possible for the teacher's announcement to be true. Where, then, is the flaw in the argument? The first stage of the argument is just:

(1) If the only afternoon examination of the week were held on Friday, then on Thursday night the student would be justified in believing that an examination will occur on Friday afternoon.

No doubt this first step of the argument looks inescapable. For, we reason, on Thursday evening the student would be justified in believing that it is now Thursday evening and an examination has not been held on this or any previous afternoon of the week; and he would also be justified in believing that an examination will be held on exactly one of the afternoon of Monday–Friday. Hence, the student would be justified in concluding, on Thursday evening, that an examination will be given on Friday afternoon.

But this reasoning depends on our ignoring part of the student's total available evidence. Grant that on Thursday night, the student remembers that the teacher is generally reliable and said:

(A) There will be an examination on exactly one of the days Monday–Friday.

(This, presumably, is the only reason *S* would have for believing that there will be an examination.) We make use of this fact about *S*'s evidence to conclude that *S* would be justified in believing (A) on Thursday night. But, in so doing, we overlook another part of the student's evidence. For he is also supposed to remember that the teacher, who is generally reliable, asserted:

(B) If an examination is held on the afternoon of day *D* then you will not be justified in believing this before that day.

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 in believing (B), for there is no epistemically relevant difference for him between the two propositions. However, if *S* were justified in believing both (A) and (B), then, realizing that it is now Thursday night and an examination has not been held on this or any previous afternoon, he would also be justified in believing:

> There will be an examination on Friday afternoon and I am not now justified in believing that there will be an examination on Friday afternoon.

Surely this is impossible. It can never be reasonable to believe a proposition of the form 'p and I am not now justified in believing p'. For if a person A is justified in believing a proposition, then he is not (epistemically) blameworthy for believing it. But if A is justified in believing that he is not justified in believing p, then he would be at fault in believing p. Hence, if A is justified in believing that he is not justified in believing p, then he is *not* justified in believing p.

The upshot is that the student would not be justified in accepting both (A) and (B). And since his evidence concerning (A) is no better than his evidence concerning (B), he would not be entitled to accept just (A).[3]

Thus we see that the seemingly air-tight argument can be stopped at the very first step, and a surprise examination is therefore possible on any day of the week. Of the premises needed for the paradox generating argument, (P_1)–(P_3) simply specify that the student's relevant intellectual abilities are unimpaired. The only substantive premises used were (P_4) and (P_5), each of which seems, *prima facie*, a highly plausible epistemic principle. This is just as it should be, for an argument which we all find seductive is not likely to be based on obviously false premises. Now, however, it can be seen that (P_5) must in fact be rejected. Even though on Thursday night the student would have good evidence for (A), etc., he would not be justified in believing (A). For he cannot be warranted in believing (A) without also being justified in accepting a proposition of the form 'p and I am not now justified in believing p'.

This solution, of couse, rests on the fact that the student has no evidence that there will be an examination other than the fact of the teacher's announcement. So one might think the paradox would break out again if we simply made a slight revision. Let us suppose the teacher makes the same announcement, but, as well, the student has strong independent evidence that an examination will be given on exactly one of the afternoons of Monday–Friday. (For example, an automatic and irreversible process has been set in motion which guarantees that an examination will be held on exactly one of the afternoons Monday–Friday.) In this revised situation, the objection to the first step of the argument is no longer open to us. But even if we therefore grant that a Friday examination would not be a surprise, we can still stop the argument at a later stage. The second step of the argument reads:

(2) If the only afternoon examination of the week were held on Thursday, then on Wednesday evening the student would be justified in believing (1), and therefore also justified in believing that an examination will be held on Thursday afternoon.

(Notice that each step of the argument after the first requires that the student be justified in believing all previous steps of the argument at the appropriate time.) Now suppose we grant that S would be justified in believing (1) on Wednesday evening. Still, we cannot reach the desired conclusion. The difficulty is that in order to be justified in believing that an examination will be held on Thursday afternoon, S would have to be justified in believing both (A) and (B); for he can rule out a Friday examination only on the basis of (B). But the student cannot be justified in believing both (A) and (B), since this would result in his being justified in believing:

> An examination will be held on Thursday afternoon and I am not now justified in believing that an examination will be held on Thursday afternoon.

Thus, under the revised conditions, the surprise examination can be held on any afternoon but the last.

In general, however we revise the situation, we must claim at *each* stage of the argument that the student is justified in using (A) to predict the date of the exam; and at *some* point we also assume that he is justified in using (B) in order to rule out certain days. But the joint use of (A) and (B) in this way is impossible, and the paradoxical argument must therefore fail.

Finally, let us take note of the starkest form of the paradox, in which the announcement is just 'There will be a surprise examination tomorrow'. My analysis has it that, even under these circumstances, an exam will be a surprise. For the student cannot be justified in believing:

> There will be an examination tomorrow and I am not now justified in believing there will be an examination tomorrow.

And he has no reason to prefer either conjunct of this conjunction. On the other hand, if we consider a revised version of this abbreviated form, the result is quite different. Suppose the student has strong independent evidence that an exam will be given. Then the student is justified in believing that an exam will be given and a surprise is not possible.

I have maintained that the prediction paradox is best seen as an epistemological problem. What we learn from the paradox is that (P_5), although *prima facie* unexeptionable, is in fact a faulty epistemic principle. Interestingly, this view of the prediction paradox is analogous to a major school of thought on the lottery paradox. Briefly, the lottery paradox involves a fair lottery in which one ticket will be drawn from a pool of, say, 1000 tickets. Let t_i be the statement that ticket$_i$ will not win. Since each of t_1, t_2, etc. is highly probable, it would appear that we are justified in believing each of:

$t_1, ..., t_{1000}$, There are 1000 tickets of which one will win.

But this set is itself inconsistent; and it entails contradictory propositions, *viz*:

Some ticket will win, No ticket will win.

A standard position on the lottery paradox is that since it cannot be rational to believe an inconsistent set of propositions; and since there is no reason to prefer some members of the set $t_1, ..., t_{1000}$, therefore we cannot believe any members of the set.[4] The analogues with the prediction paradox are fairly obvious. In each case, what we have to learn from the paradox is that (P_5), although *prima facie* sound, is not an acceptable epistemic principle. Strong support by the total available evidence may not confer the right to believe.

An alternative approach to the lottery paradox has it that while we cannot admit that it is reasonable to accept contradictory beliefs (p, not-p), we can tolerate an inconsistent set of justified beliefs; and we can rule out justified belief in contradictory propositions by denying (P_4). Perhaps those who can countenance an inconsistent set of beliefs would also have no trouble accepting:

p, I am not justified in believing p.

In any case, it is generally acknowledged that some version of (P_4) is essential to the satisfactory progress of our reasoning. We do not want to rule out all inferences from multiple premises. At most, we want to revise (P_4); and an obvious way to do so is to add the additional requirement that the conjunction of the premises $p_1, ..., p_n$ must have a certain minimum probability. The reader who has qualms concerning (P_4) can satisfy himself that any reliance on (P_4) in this paper does not seem to violate this additional requirement.

That is, it would appear that the use of (P_4) has been quite innocuous.

To sum up: The chief lesson the prediction paradox teaches us is that (P_5) must be rejected. Hence, if my view of the lottery paradox is correct, it is a close cousin of the prediction paradox. Each teaches us essentially the same lesson: roughly put, that good evidence is not sufficient for justified belief.

This analysis of the prediction paradox, and the rejection of (P_5), rest of course on two further assumptions:

(I) It is impossible to be justified in believing a pair of propositions of the form 'p, I am not justified in believing p'.

(II) If it is impossible to be justified in believing each member of the set $p_1, ..., p_n$; and there is no proper subset of $p_1, ..., p_n$ of which this is true; and you have equally good reason to believe each of $p_1, ..., p_n$, then you are not justified in believing any one of these propositions.

The reasoning underlying (I) has been indicated earlier in this paper. Assumption (II), which asserts the non-arbitrary nature of justified belief, is also essential to the analysis of the prediction paradox. Without it, we cannot ensure that S is not justified, on Thursday night, in expecting an examination the next day. Notice that (II) is also relevant to the lottery paradox. A possible line on the lottery problem is that one may avoid inconsistency by believing just a subset of $t_1, ..., t_{1000}$. Which subset is believed, it might be suggested, is a matter of indifference, as long as the conjunction of propositions believed has a certain minimum probability. The effect of (II) is to rule out this possible approach.

My analysis rests, then, on these two epistemological premises. If they are as plausible as I think, then those of us who are not students may cease worrying about surprise examinations.

NOTES

[1] For an instance of what appears to be similar reasoning, see G. C. Nerlich: 1961. 'Unexpected examinations and unprovable statements', Mind 70, p. 507. It should also be remarked that to condone acceptance of the student's argument as part of one's 'solution' will not resolve the problem.

[2] Some examples of this approach are: Brian Medlin: 1964, 'The unexpected examination', American Philosophical Quarterly 1, pp. 66–72; Peter Y. Windt: 1973, 'The liar

in the prediction paradox', American Philosophical Quarterly 10, pp. 65–68; Martin Edman: 1974, 'The prediction paradox', Theoria 40, pp. 166–175. It should be noted that the versions of (T) suggested as correct interpretations are quite unnatural as readings of the announcement.

[3] A consequence of this argument, roughly put, is that while the announcement may be true, it cannot be accepted. Others whose views can be (roughly) categorized in this way are: R. Binkley: 1968, 'The surprise examination in modal logic', Journal of Philosophy 65, pp. 127–136, and C. Wright and A. Sudbury: 1977, 'The paradox of the unexpected examination', Australasian Journal of Philosophy 55, pp. 41–58. The philosophical underpinings, however, are quite different. In my view, each of these analyses utilizes dubious or, at least, highly controversial assumptions.

[4] Although I share this view of the lottery paradox, a defence of it is beyond the scope of the present paper.

Glendon College,
York University,
2275 Bayview Avenue,
Toronto,
Ontario, M4N 3M6,
Canada