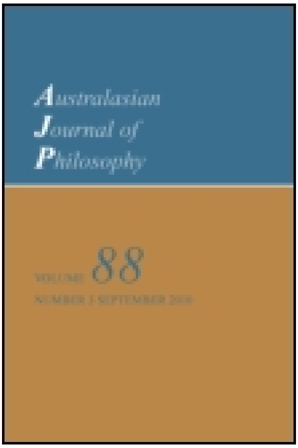

Australasian Journal of Philosophy
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/rajp20

The prediction paradox: Resolving recalcitrant variations

Doris Olin [a]
[a] York University

Available online: 02 Jun 2006

Australasian Journal of Philosophy
Vol. 64, No. 2; June 1986

THE PREDICTION PARADOX: RESOLVING RECALCITRANT VARIATIONS

Doris Olin

Three ingenious variations of the traditional prediction paradox have recently come to light in a paper by Roy A. Sorensen.[1] The variations demonstrate that certain standard features of the familiar version of the paradox are not essential. For example, the paradox need not even involve a prediction. Sorensen also argues that none of these new versions can be handled by any of the analyses offered to date and, hence, we do not yet have a comprehensive solution to the paradox.

In a recent paper, I maintain that the traditional paradox is an epistemic problem and that it rests on a certain initially plausible, but mistaken, epistemic principle.[2] I shall argue here that, despite the cleverness of the new variations, they fall to the same solution and, thus, do not call for any further 'solutions'.

First, let me outline my analysis of the paradox. The familiar version of the prediction paradox involves a teacher who announces to her student that she will give exactly one surpise examination next week. The student objects that the announcement cannot be satisfied, reasoning as follows: 'If the examination were held on Friday, then on Thursday evening, realising that no examination had yet been given, I would reasonably expect it on Friday; hence a Friday examination would not be a surprise. But, if the exam were given on Thursday, then on Wednesday night I would be aware that an examination had not yet been given and, recognising that it cannot be given Friday, would expect it on Thursday; so a Thursday examination would not be a surprise. And so on through the remaining days. Consequently, the surprise examination cannot be given.' But, of course, the surprise examination can be given—on Wednesday, for example. So we have a paradox. A seemingly impeccable piece of reasoning leads to a conclusion which is clearly false. To dispell the paradox, it is necessary to isolate the flaw in the student's argument.

Whether or not the examintion will be a surprise turns on what the student is justified 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, prior to the day of the examination, that it will occur on that day. With this interpretation, let us look more closely at the paradox-generating argument. The first step of the argument is:

[1] Roy A. Sorensen, 'Recalcitrant Variations of the Prediction Paradox', *Australasian Journal of Philosophy* 60 (1982), pp. 355-362.
[2] The analysis is presented in 'The Prediction Paradox Resolved', *Philosophical Studies* 44 (1983), pp. 225-234.

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

No doubt this step of the argument looks inescapable. For we assume the student's logical abilities and memory are beyond reproach, the teacher is generally reliable, and so on. But let us look more closely. The teacher's announcement can be rendered in terms of two propositions:

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

(B) If an examination is held on a given day, you will not be justified in believing this before that day.

A crucial assumption underlying (1) is that the student would be justified in believing (A) on Thursday night, since he remembers what the teacher said and knows that the teacher is reliable. But this assumption is mistaken. Consider. If the student is justified in believing (A) on Thursday night, then he is also justified in believing (B), for there is no epistemically relevant difference for him between the two. (Since both are based just on the teacher's testimony, his evidence for one is as good as his evidence for the other.) But if he is justified in believing *both* (A) and (B), then he is also justified in believing:

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

Surely, however, one can never be justified in believing a proposition of the form '*p* and I am not now justified in believing *p*'. Hence the student cannot justifiably believe the teacher's announcement on Thursday night and the argument is blocked at the very first step. (Notice that this reasoning shows only that the student cannot justifiably believe the announcement on Thursday night, given no previous examination. There is no reason to suppose that he cannot justifiably believe it earlier.)

The analysis, of course, depends on the fact that, in the familiar version, the student has exactly the same evidence for (A) and (B). Might not the paradox break out again if we suppose that, as well as the announcement, the student has strong independent evidence for (A)? In the revised version, where there is better evidence for (A), the objection to the first step is no longer open to us. Even if the only examination were held on Friday, on Thursday evening the student would still be justified in believing (A), for he would have reason to prefer it to (B). So a Friday examination would not be a surprise. But we can still stop the argument at a later stage. Consider the second step:

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

Suppose that the student were justified in believing (1). Still, in order to infer that an examintion will be held on Thursday, the student would have to be justified in accepting both (A) and (B); for he can rule out a Friday examination only on the basis of (B). But if this were so, the student would be entitled to believe:

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

So, in the revised version, the argument fails at step (2), and the surprise examination can be held on any day but the last. In general, when we revise the basic situation, we find that at *each* stage of the argument the subject must be able to rely on (A) to predict the date of the examination; and at *some* stage he must be justified in using (B) to rule out certain days. But the *joint* use of the two propositions is impossible and the argument must, therefore, fail.

This, in essence, is my treatment of the paradox. A fuller account would refer explicitly to the premises on which the paradoxical argument depends. One such premise is:

(P) If S is justified in believing $p_1, \ldots, p_n$, $p_1, \ldots, p_n$ strongly confirm q, S sees this and has no other evidence relevant to q, then S is justified in believing q.

(P) is, of course, a highly intuitive epistemic principle. This is as it should be, for an argument which we all find seductive is not likely to be based on obviously false premises. Nevertheless, the chief lesson of the prediction paradox is that (P) is not an adequate epistemic principle. Even though on Thursday night, the student has good evidence for (A), recognizes the force of the evidence, and so on, he is not justified in believing (A). The principles of epistemology turn out to be even more complex than we had originally thought.

This analysis of the paradox, and the rejection of (P), rely heavily on the following principle:

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

The reasoning underlying (I) may be put thus. If a person A is justified in believing *p*, then he is not (epistemically) blameworthy in believing *p*. 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*, this precludes his being justified in believing *p*. Notice that although (I) guarantees that Jane cannot justifiably believe both that *p* is true and that she is not justified in believing *p*, it may be perfectly permissible for Jake to believe that *p* is true and that Jane is not justified in believing *p*. Thus, one consequence of (I) is that what one may believe depends, in part, on who one is.

Can this analysis of the traditional version now be applied to Sorensen's variations? To do so, we must construe each of the three situations in terms of a proposition analogous to (A), which describes the basic set-up, and one analogous to (B), which states that the outcome will be a surprise. The notion of a surprise will be interpreted in terms of the absence of *justified* belief prior to a certain time. There will also be a general assumption made to the effect that there is no relevant epistemic difference between the subject's evidence for the proposition corresponding to (A), and his evidence for the proposition corresponding to (B); I have indicated above how the analysis may be applied when there is such a difference.

I. *The Designated Student Paradox*

Of five students, Art, Bob, Carl, Don and Eric, one is to be given an examination. The students are lined up in alphabetical order, so that each can see the backs of those before him. The teacher has four silver stars and one gold star which will be placed on the students' backs; the gold star designates the student who will be examined. Proposition (A1) gives just this information; while proposition (B1) states that the designated student will not be entitled to believe he is the designated student until after the students break formation.

Having been informed of (A1) and (B1), the students generate the following argument: If Eric were the designated student, then he would see four silver stars ahead of him and would thus be able to infer that he is the designated student; so he is not the designated student. But if Don were the designated student, then since he would see three silver stars in front of him, and would realise that Eric is not the designated student, he would be entitled to believe that he is the designated student; so he is not the designated student. And so on. The students conclude that the specified examination cannot be given. They then break formation, and Carl is surprised to learn that he is the designated student.

As with the original version, the argument falters in the very first step:

(1) If Eric were the designated student, then he would be justified in believing this before breaking formation.

Eric's belief would, presumably, be based on (A1). But we assume that (A1) and (B1) are epistemically indistinguishable for him—he has exactly the same evidence for each. So he could believe (A1) only if he were entitled to accept both (A1) and (B1). But, in that case, he would also be justified in believing:

> I am the designated student and I am not now justified in believing that I am the designated student.

This is surely impossible. The upshot is that *Eric* cannot believe the teacher's announcement, *if* he is in fact the designated student and sees four silver stars. (The other students are, of course, entirely justified in accepting the announcement, no matter who is the designated student.) Thus, the argument collapses and the paradox is resolved.

II. *The Paradox of the Undiscoverable Position*

This version is intended to show that a single, rigid order of elimination is not an essential feature of the prediction paradox. Proposition (A2) gives the following information concerning a certain game. The game utilises the matrix below.

1	2	3
4	5	6
7	8	9

The player is assigned an initial position in the matrix, and the object of the game is to discover this position. The player may move Up, Down, Left or Right, one box at a time, and is allowed at most two moves. If she bumps into a wall, say by moving Left, then her position is unchanged and her move is recorded as $\overline{L}$.

The 'examiner' informs the player of (A2), and also of:

(B2) You are in an undiscoverable position.

The player then declines any moves and offers the following reasoning: 'If I were in one of the corners, then there would be a sequence of two moves which would reveal my position. For example, if I were in 1, then if I made the sequence of moves $\overline{L}$, $\overline{U}$, I would discover my position. So I am not in one of the corners. But given this, I can also eliminate 2, 4, 6 and 8, since any bump from a first move will show me where I am. The only remaining possibility is 5; so if I were in 5, my position would clearly be discoverable.[3] Thus, I cannot, in this game, be put in an undiscoverable position.'

Obviously, the player can be put in an undiscoverable position if she is not informed of (B2); nor does it seem that this information should make a relevant difference. Again, we must look for the flaw in the reasoning. The first step of the argument is:

(1) If the player were in one of the corners, then she would be in a discoverable position.

The objection to (1) is no longer difficult to discern. Suppose the player is in 3, and does make the sequence of moves $\overline{U}$, $\overline{R}$. From (A2) and (B2), the player might then infer:

I am in 3 and I am not now justified in believing that I am in 3.

But one cannot justifiably believe such a proposition. And since there is no relevant epistemic difference between (A2) and (B2), the player cannot accept either. (The announcement, notice, becomes incredible only when the player is in a corner and *has made* certain moves. Otherwise, there is no reason to suppose that (A2) and (B2) cannot be believed.) The upshot is that even if the player were in 3 and did make the approriate moves, she would not

[3] Sorensen points out that the argument could also be developed in such a way as to make *any* position the last remaining position. I might, for instance, eliminate the corners, then 2, 6, 8. This leaves 4 and 5. But 5 could be discovered by the moves R, $\overline{R}$; thus 4 is the only remaining possibility.

be able to infer her position. So the corners cannot be eliminated and the paradoxical argument does not go through.

III. *The Sacrificial Virgin Paradox*

The final version of the paradox is intended to show that the paradox does not presuppose knowledge of the number of alternatives. Proposition (A3) contains the following information. The inhabitants of a tropical paradise plan to sacrifice a virgin to the local volcano. A number of virgins are blindfolded and brought before the volcano. They all hold hands in a line and can only communicate the sentence: 'No one to your right is a sacrificial virgin.' This is done by squeezing the hand of the virgin to one's left. The virgins are chosen for their reliability and logical abilities, and are instructed to give the signal if and only if they know the truth of what is communicated. The chief takes the leftmost virgin to the mouth of the volcano and, if the offering is acceptable, sacrifices her and sends the others home. If not, he tries again with the new leftmost virgin. As well as (A3), the virgins are informed of:

(B3) The sacrificial virgin will not be justified in believing, prior to being tossed in, that she is the sacrifice.

The argument which purports to show that the ceremony cannot take place as described is as follows. Any virgin is either (a) the rightmost (b) a middle virgin or (c) the leftmost. (a) If the sacrificial virgin is the rightmost, then she realises she is the rightmost (since her right hand is free), and thus goes to the volcano aware that she is the last alternative, and thus is able to infer that she is the sacrifice. So she cannot be the sacrificial virgin. The rightmost virgin will, given her logical powers, see this and, given her reliability, will signal by squeezing the hand of the virgin on her left, who is either a middle or the leftmost virgin. (b) If she is a middle virgin, she will see that, if she is to be offered, she will be aware beforehand that no one to her left has been sacrificed. And, since she has received the signal on her right hand, she will be entitled to infer that she is the sacrifice. So, she realises, she cannot be the sacrifice, and she squeezes the hand on her left. Similarly for all the middle virgins. (c) If the sacrificial virgin is the leftmost then, since she has received the signal, she realises that she is the only remaining virgin and is therefore the sacrifice. So she cannot be the sacrifice and the ceremony is impossible.

In the by now familiar manner, we isolate the first step in the reasoning:

(1) If the rightmost virgin is the sacrifice, then she will be justified in believing this before being tossed in.

Consider the position of the virgin once both her hands are free. She has been informed of both (A3) and (B3). This information, plus her knowledge that both her hands are free, implies:

I am the sacrificial virgin and I am not now justified in believing that I am the sacrificial virgin.

Since she cannot believe this, she cannot believe both (A3) and (B3) once both her hands are free. And given that there is no reason to prefer either, she cannot believe (A3). Hence, she is not entitled to believe that she is the sacrificial virgin and the paradox is dispelled.

It is worth remarking that, in each of the versions we have been discussing, the subjects must in fact appeal to premises other than just the analogues of (A) and (B) in order to infer a proposition of the form 'p and I am not now justified in believing p'. For example, in the designated student variation, Eric must rely on the fact that each student ahead of him has a silver star. In discussing this variation, I assumed, tacitly, that a proposition based on present perception is clearly to be preferred to either (A1) or (B1). More explicitly, the essential structure of the paradoxical situation may now be described as follows. In each step of the paradox generating argument, we envision a case in which there is strong support by the total evidence for each member of a set of propositions $\{p_1, \ldots, p_n\}$; but this set jointly implies something which one cannot be justified in believing. If there is no epistemically relevant way to discriminate among members of the set then, I would claim, one should not believe any member of the set. But suppose one can discriminate. Perhaps p_1 is based on direct perception, while the other members of the set are based on the evidence of testimony. Assuming that this gives p_1 preferred status, this proposition should be included in the corpus of one's belief. If one can discriminate among the remaining members of the set, then some further propositions may also be accepted. And so on, until there is no longer ground for preference.

An interesting variant of the above is the case in which the set $\{p_1, \ldots, p_n\}$ implies a contradiction. That this can happen, *i.e.* that a set of propositions, each member of which is strongly confirmed by the total evidence, may be inconsistent is clearly demonstrated by the lottery paradox. Consider a 1,000 ticket fair lottery with exactly one winner. Each member of the set $\{t_1, \ldots, t_n\}$ (where t_1 = Ticket 1 will not win) is strongly supported by the total evidence. But the totality of such propositions, together with the proposition that exactly one of 1,000 tickets will win, form an inconsistent set and thus imply a contradiction. In my view, a proper treatment of this paradox will bear a close resemblance to the analysis of the prediction paradox, in that each requires, for resolution, a denial of (P). But that is another story which is beyond the scope of this paper.

In a second, yet more recent paper, Sorensen has gone on to provide his own solution to the prediction paradox.[4] I should like to conclude by contrasting my analysis with the one he proposes.

Sorensen interprets the notion of surprise in terms of lack of *knowledge* prior to a certain time. To say the examination will be a surprise, on Sorensen's interpretation, is to say that the student will not *know* beforehand when it will occur. Now consider the position of the student of Thursday night when no examination has yet been given. Suppose that he knows the truth of the

[4] 'Conditional Blindspots and the Knowledge Squeeze', *Australasian Journal of Philosophy* 62 (1984), pp. 126-135. The foregoing was written without the benefit of this yet more recent paper by Sorensen. I should like to thank a referee of this journal for bringing it to my attention.

announcement at this time. This would mean that he *knows* both that (a) there will be an examination on Friday and that (b) he does not now know that there will be an examination on Friday. But this is a logical impossibility. Where 'K*ap*' signifies that *a* knows that *p*, a proposition of the form

$$Ka(p \cdot \sim Kap)$$

is, Sorensen points out, logically inconsistent and implies a contradiction. Hence the student cannot know the announcement on Thursday night.

The recognition of this logical inconsistency is at the core of Sorensen's analysis. The inconsistency stems from certain logical features of 'knows'—in particular, the fact that 'knows' is a success term, that is, 'K*ap*' entails the truth of *p*. Hence, if 'K*a*(~ K*ap*)' is true, then it follows that ' ~ K*ap*' is true as well. This feature of 'knows' is not particularly related to its being an epistemic term; it is shared, for example, by 'brings about'.

Although I cannot accept some of the details of Sorensen's account, there can be no doubt of the inconsistency of 'K*a*(*p*• ~ K*ap*)'. Nevertheless, I do not think that recognition of this logical feature of 'knows' can completely dispell the prediction paradox, however one makes use of it.

I have, of course, interpreted 'surprise' in terms of the absence of justified belief, and I take this interpretation to be preferable. Insofar as the notion of surprise is connected with that of knowledge, it is, I think, through the normative component of knowledge—i.e. justification. But let us leave aside the issue of which construal is preferable. What is clear is that even if we were to deal, in Sorensen's way, with the prediction paradox formulated in terms of 'knows', we could *go on* to spell out a version of the paradox generating argument in which 'surprise' is construed in terms of the absence of justified belief. This argument is, as we have seen, highly seductive; and nothing that Sorensen has said about the logical features of 'knows' can help us to refute it. To rebut step (1) of this argument, as formulated in my analysis above, we require something more. What we require is a substantive normative epistemic principle. That is, we must invoke principle (I) and provide a defence of it. To sum up: We see that central to Sorensen's approach is the observation that 'knows' is a success term, and, as a result, certain knowledge claims are logically inconsistent; while my analysis rests essentially on a substantive epistemic principle about justified belief. What I have argued is that this epistemological machinery is necessary to completely dispell the paradox.

It is worth remarking that even Sorensen's version of the paradox expressed in terms of 'knows' is not completely resolved without commitment to a principle similar to (I). Grant that on Thursday night, the student cannot *know* the truth of the announcement (on pain of contradiction). Nevertheless, one might still raise the question whether he is *justified* in believing the announcement. If he is, then he is justified in believing that there will be an examination tomorrow. If the examination will in fact be given on Friday, then he has a *true* justified belief that there will be an examination tomorrow and therefore (assuming the traditional analysis of knowledge) knows this. Hence a Friday examination will not be a surprise.

In order to rule out this reasoning, and the possibility of justified belief in the announcement on Thursday night, one must have recourse to a principle which ensures that, even if one has good evidence, one cannot be justified in believing a proposition of the form '*p* and I do not now know that *p*'. Hence, it would seem that a complete resolution of even Sorensen's version of the paradox will involve commitment to a substantive epistemic principle analogous to (I). In any of the forms considered, then, resolution of the prediction paradox appears to require an epistemic principle.

One motivation for wanting to solve a paradox is that one may thereby learn something of philosophical value. What we learn from my analysis, and the excursion into epistemology, I maintain, is that principle (P) is false. Strong support by the total available evidence is not sufficient for justified belief. A familiar rule of acceptance, much discussed in epistemology and philosophy of science, is the Rule of High Probability. (Accept a proposition if it is highly probable on the total evidence.) If my analysis is correct, then we must bear it in mind when we come to think about the Rule of High Probability and the lottery paradox to which this Rule gives rise.

York University Received December 1984

[5] I am grateful to Bernard D. Katz and a referee of this journal for helpful comments.