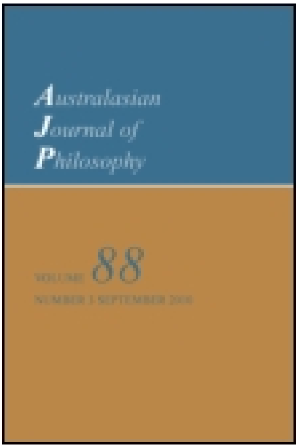

Australasian Journal of Philosophy
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/rajp20

Recalcitrant variations of the prediction paradox

Roy A. Sorensen [a]
[a] Michigan State University, East Lansing

Available online: 02 Jun 2006

To cite this article: Roy A. Sorensen (1982): Recalcitrant variations of the prediction paradox, Australasian Journal of Philosophy, 60:4, 355-362

To link to this article: http://dx.doi.org/10.1080/00048408212340761

Australasian Journal of Philosophy
Vol. 69, No. 4; December 1982

RECALCITRANT VARIATIONS OF THE PREDICTION PARADOX

Roy A. Sorensen

After informal solutions to the prediction paradox failed, philosophers adopted the formal approach. Here one rigorously reconstructs the argument in order to lay bare its essential premises. One then solves the paradox by providing a good argument for rejecting at least one of these premises. The chief hazards of this approach are the failure to faithfully represent the paradox and the failure to provide a comprehensive solution (one which solves all the variations of the paradox). I think all proposals in the self-referential tradition exemplify the first failure.[1] However, I shall concentrate on the second failure. I shall defend the thesis that the structure of the prediction paradox has not yet been formulated by introducing three variations of the paradox which do not conform to any of the formulations present in the voluminous literature on this topic. Special attention shall be given to how two recent proposals are refuted by a variation I call the designated student paradox.

One familiar version of the prediction paradox involves a teacher who announces to his students that he will give exactly one surprise examination next week. A clever student objects that such an examination is impossible. The examination cannot be given Friday since, after Thursday's class, the students would know that the examination must be on Friday. The examination cannot be given Thursday since, after Wednesday, the students would know that the examination must be on Thursday (having ruled out Friday). In a like manner, the student eliminates Wednesday, Tuesday, and Monday, thereby 'proving' that the examination cannot be given. Yet, if the teacher gives the examination, say, Wednesday, it will be a surprise, and thus the announcement will be true.

Although commentators are quick to generalise the paradox to the n-day case, none to my knowledge, consider the $1 \leqslant m \leqslant n$ where n is the number of days and m is the number of surprises. It is easy enough for the students to argue that the m^{th} examination cannot be a surprise, since after the $m - 1$ surprise they can perform the standard elimination for the single surprise case. The students might be satisfied with merely showing that not all of the examinations can be surprises or they might become more ambitious and try to

* I thank Herbert Hendry and Michael Pokriefka for their many valuable comments on an earlier draft.

[1] Assimilating the prediction paradox to the self-referential paradoxes was first suggested by R. Shaw in 'The Paradox of the Unexpected Examination' *Mind* (July, 1958).

eliminate the rest. They could argue that the m – 1th examination cannot be a surprise since after the m – 2nd surprise, there would be only one surprise examination forthcoming (since the mth examination has been shown to be unsurprising), thus enabling them to employ the standard elimination argument once more. Having tamed the last two examinations the students will run the rest through the routine. The upshot seems to be that it is impossible to inform someone that he will be surprised m times within a period of n occasions.

In 'The Paradox of the Unexpected Examination', Crispin Wright and Aidan Sudbury list six conditions which would make a solution to the paradox intuitively satisfying.[2] Among other things, the solution should make it clear that the announcement can be fulfilled, do justice to the intuitive meaning of the announcement, and allow that the students be informed by the announcement. The authors argue that their conditions can be met by rejecting what others have called the temporal retention principle or the memory principle. Roughly, the temporal retention principle states that if one knows or reasonably believes *p* at time *t*, then one continues to do so after *t*. This principle has been standardly invoked by other commentators who readily concede that real-life students forget, go insane, die, etc. These commentators maintain that nevertheless, one can stipulate these calamities away to expose a genuine paradox involving ideal thinkers. As Hintikka's *Knowledge and Belief* shows, this sort of knowledge and belief, though not intended to be psychologically accurate, is philosophically interesting insofar as it describes how knowledge and belief claims commit us to other claims. So commentators stipulate that the teacher is known to be completely trustworthy; that death, insanity, and forgetfulness do not victimise the parties involved, that the parties each know this and know all involved know, etc. However, Wright and Sudbury do not base their rejection on such considerations. Even ideal thinkers in a favourable epistemological environment would be forced to change their minds. Wright and Sudbury point out that the teacher's announcement resembles the Moorean sentence p & $-Bap$. Person a can consistently believe that it was the case that p & $-Bap$ or that it will be the case that p & $-Bap$, but he cannot believe that it is now the case that p & $-Bap$.[3] Moorean sentences are usually asymmetrically credible. Person b can now believe p & $-Bap$ because the Moorean sentence is not about b. By announcing a surprise examination, the teacher is in effect informing each of the students that a Moorean sentence about him will be made true by Friday. If the examination is not given by Thursday, the announcement combined with the student's knowledge that it has not yet been given makes the announcement tantamount to an unbelievable present tense Moorean sentence. Thus Wright and Sudbury can maintain that the announcement did inform the students and yet, on pain of inconsistency, the students can no longer know the

[2] Crispin Wright and Aidan Sudbury 'The Paradox of the Unexpected Examination' *Australasian Journal of Philosophy* (May, 1977) see p. 42 for the list.

[3] There are many analyses of Moorean sentences. The most widely known is in Jaakko Hintikka's *Knowledge and Belief* (Ithaca: Cornell University Press, 1962).

announcement is true if the examination is not given by Thursday. It also follows that since the announcement only states that the students will be surprised, it is possible for a nonstudent to know that the examination will be given Friday even though he has the same evidence as the students. A startling corollary of this solution (though the authors fail to state it) is that ideal thinkers can disagree simply by virtue of their identities.[4]

Although I endorse the rejection of the temporal retention principle, I do not believe one thereby solves the prediction paradox. Consider the designated student paradox. Here, only one examination is to be given to one of five students: Art, Bob, Carl, Don, Eric. The teacher lines them up alphabetically so that Eric can see the backs of each of the four students in front of him, Don can see the backs of the three students in front of him (but not Eric's since Eric stands behind him), and so on. The students are then shown four silver stars and one gold star. One star is put on the back of each student. The teacher then announces that the gold star is on the back of the designated student. He informs them that the designated student must take the examination. The examination is unexpected in the sense that the designated student will not know he is the designated student until after the students break formation. One of the students objects that the examination is impossible. 'We all know that Eric is not the designated student since, if he were, he would see four silver stars in front of him and deduce that he must have had the gold star on his back. But then he would know that he was the designated student. The designated student cannot know he is the designated student; contradiction. We all know that Don cannot be the designated student since, if he were, he would see three silver stars in front of him, and since he knows by the previous deduction that Eric has the remaining silver star, he would be able to deduce that he is the designated student. In a similar manner, Carl, Don, and Art can be eliminated. Therefore, the examination is impossible.' The teacher smiles, has them break formation, and Carl is surprised to learn that he has the gold star, and so is the designated student, and so must take the examination.

In this variation, knowledge is accumulated perceptually rather than by memory. One might object that this paradox relies on the temporal retention principle since the students must continue to know the announcement as they perform the deduction. Even if one is a quick deducer, a small amount of time is needed. One might reply that ideal thinkers could perform the deduction instantaneously so that the paradox holds for ideal thinkers. The objector might then counter that ideal thinkers cannot perform instantaneous deductions. Since I am not sure how such an issue can be settled, I shall assume for the sake of argument that the temporal retention principle is needed to insure a few moments of knowledge. But this should give Wright and Sudbury little comfort since their claim that the announcement is informative is then dubious. They

[4] In 'Disagreement Amongst Ideal Thinkers' (forthcoming in *Ratio*, probably December 1981), I argue for this point without mentioning the prediction paradox.

would have to maintain that the students were informed only for a moment. Suppose Eric has the gold star on his back. He knows this before the announcement is made by elimination since he sees four silver stars in front of him. Eric is then informed that the person with the gold star is the designated student and that the designated student does not know he is the designated student. Whether or not Eric infers that he is the designated student but does not know it, he is inconsistent. Since one cannot be informed in this manner and remain consistent, it follows that an ideal thinker cannot be informed in this manner. Therefore, Wright and Sudbury could no longer meet their condition of informativeness.

The designated student paradox also undermines proposals made by J. McLelland, C. Chihara, and Craig Harrison.[5] They show that the traditional versions of the prediction paradox rely on the KK principle: if one knows that p, then one knows that one knows that p. They note that even prior to consideration of the prediction paradox, the KK principle is controversial. Since they deem this principle as the most dubious of the alternatives, and since rejecting it yields their desired consequence that the examination can be given on any day except the last, they conclude that the prediction paradox supports the rejection of this principle.

However, the designated student paradox does not rely on the KK principle. To show why, I shall first formalise the content of the announcement. Let 'p_1' stand for 'Art is the designated student', 'p_2' for 'Bob is the designated student', and so on. The first five letters of the alphabet denote Art, Bob, Carl, Don, and Eric. Thus one reads '$\mathrm{K}ap_1$' as 'Art knows that Art is the designated student'.

(1) $[(p_1 \supset -\mathrm{K}ap_1) \,\&\, (p_2 \supset (-\mathrm{K}bp_2 \,\&\, \mathrm{K}b{-}p_1)) \,\&\, \ldots \,\&\, (p_5 \supset (-\mathrm{K}ep_5 \,\&\, \mathrm{K}e{-}(p_1 \vee p_2 \vee p_3 \vee p_4))) \,\&\, (p_1 \vee p_2 \vee p_3 \vee p_4 \vee p_5)]$

One justifies the accumulation of $\mathrm{K}{-}p$'s in (1) by the increasing perceptual knowledge enjoyed by those behind Art. In order to meet the requirement of informativeness, one is tempted to prefix (1) with $(x)\mathrm{K}x$ where x ranges over the students. However, $(x)\mathrm{K}x(1)$ corresponds to a variation where the teacher makes a private announcement to each student (so that none know the others know the announcement). Supposing p_5 and universally instantiating to e would establish $\mathrm{K}e{-}p_5$ but not $\mathrm{K}d{-}p_5$. In the case where everyone knows (1) but no one knows that everyone knows (1), Don does not know that Eric knows (1). For all Don knows Eric may be ignorant of the announcement and therefore the announcement could be fulfilled by making Eric the designated student. In the original version, however, the announcement is addressed to the class as a whole, and so it appears that a faithful representation demands $(x)\mathrm{K}x(y)\mathrm{K}y(1)$. If everyone knows that everyone knows that (1), then Don does indeed know that Eric knows that Eric is not the designated student.

[5] See J. McLelland's and C. Chihara's 'The Surprise Examination Paradox' *Journal of Philosophical Logic* (February, 1975) and Craig Harrison's 'The Unanticipated Examination in View of Kripke's Semantics for Modal Logic' *Philosophical Logic* ed. Davis et. al.

For the sake of brevity, a contradiction shall be deduced from the two person version. The full content of the announcement corresponding to this case of the paradox is:

(2) $(p_1 \supset -\mathrm{K}ap_1)$ & $(p_2 \supset (-\mathrm{K}bp_2$ & $\mathrm{K}b{-}p_1))$ & $(p_1 \vee p_2)$

The inconsistency of K*a*K*b*(2) can be established in an epistemic version of KT.[6] Any normal modal analysis of epistemic logic must contain KT, so acceptance of any representation of the situation which implies K*a*K*b*(2) requires rejection of this kind of analysis. In addition to the standard sentenial rules of inference (TF), I appeal to the following rules:

KD: $\dfrac{\mathrm{K(A \;\&\; B)}}{\mathrm{KA \;\&\; KB}}$ and $\dfrac{\mathrm{K(A \supset B)}}{\mathrm{KA \supset KB}}$ KI: $\dfrac{\vdash \mathrm{A}}{\mathrm{KA}}$ KE: $\dfrac{\mathrm{KA}}{\mathrm{A}}$

KEI: $\dfrac{\begin{array}{c}\mathrm{KA}\\ \mathrm{KB}\end{array}}{\mathrm{KC}}$ where (A & B) truth-functionally imples C

KD entitles distribution of the knowledge operator over conjunction and material conditionals. KI makes all logical truths known and KE represents 'Knowledge implies truth'. KEI insures that the knower knows all the consequences of what he knows, and can be derived from the other rules.

{1}	1. $\mathrm{K}b$ (2)	Assumption
{2}	2. p_2	Assumption
{1}	3. $\mathrm{K}b(p_1 \vee p_2)$	1, KD, TF
{1}	4. (2)	1, KE
{1, 2}	5. $-\mathrm{K}bp_2$ & $\mathrm{K}b{-}p_1$	2, 4, TF
{1, 2}	6. $\mathrm{K}bp_2$	3, 5, TF, KEI
{1, 2}	7. $-\mathrm{K}bp_2$	5, TF
{1}	8. $-p_2$	2, 6, 7, Reductio
{ }	9. $\mathrm{K}b(2) \supset -p_2$	1, 8, Conditionalisation
{ }	10. $\mathrm{K}a(\mathrm{K}b(2) \supset -p_2)$	9, KI
{ }	11. $\mathrm{K}a\mathrm{K}b(2) \supset \mathrm{K}a{-}p_2$	10, KD
{12}	12. $\mathrm{K}a\mathrm{K}b(2)$	Assumption
{12}	13. $\mathrm{K}a{-}p_2$	11, 12, TF
{ }	14. $\mathrm{K}b(2) \supset (2)$	1, 4, Conditionalisation
{ }	15. $\mathrm{K}a(\mathrm{K}b(2) \supset (2))$	14, KI
{ }	16. $\mathrm{K}a\mathrm{K}b(2) \supset \mathrm{K}a(2)$	15, KD
{12}	17. $\mathrm{K}a(2)$	12, 16, TF
{12}	18. $\mathrm{K}a(p_1 \vee p_2)$	17, KD, TF
{12}	19. $\mathrm{K}ap_1$	13, 18, KEI

[6] For an explanation of KT see Brian Chellas's *Modal Logic* (New York: Cambridge University Press, 1980).

{12}	20. $p_l \supset -\mathrm{K}ap_l$	17, KE, TF
{12}	21. p_l	19, KE
{12}	22. $\mathrm{K}ap_l \,\&\, -\mathrm{K}ap_l$	19, 20, 21, TF

Since $\mathrm{K}a\mathrm{K}b(2)$ is inconsistent in KT, anything implying it is likewise inconsistent. Thus the unquantified $\mathrm{K}a\mathrm{K}b(2)$ & $\mathrm{K}b\mathrm{K}a(2)$ cannot be a faithful representation of a public, informative announcement to the pair. One can infer $\mathrm{K}a\mathrm{K}b(2)$ from $(x)\mathrm{K}x(y)\mathrm{K}y(2)$ in a quantified KT, so it is inconsistent as well.

The traditional prediction paradoxes as well as the designated student paradox involve a single, rigid order of elimination. The paradox of the undiscoverable position is intended to show that this feature is not essential to the prediction paradox. Consider the following game played in the matrix below.

1	2	3
4	5	6
7	8	9

The object of the game is to discover where you have been initially placed. The seeker may only move *U*p, *D*own, *L*eft, or *R*ight, one box at a time. The outer edges are called walls. If the seeker bumps into a wall, say by moving left from 1, his move is recorded as $\bar{\mathrm{L}}$, and his position is unchanged. Bumps help the seeker discover his initial position. For instance, if he is at 7 and moves U, U, $\bar{\mathrm{L}}$, the seeker can deduce that he must have started from 7. The seeker has discovered where he started from if he obtains a completely disambiguating sequence of moves.

If the seeker is given only two moves, it is possible to put him in an undiscoverable position. For instance, if he is put in position 4, every possible two move sequence is compatible with him having started from some other position. Now suppose the seeker is told 'You have been put in an undiscoverable position'. He disagrees and offers the following reductio ad absurdum:

> Suppose I am in an undiscoverable position. It follows that I cannot be in any of the corners since each has a completely disambiguating sequence. For instance, if I am in 3, I might move $\bar{\mathrm{U}}$, $\bar{\mathrm{R}}$, and thereby deduce my position. Having eliminated the corners, I can also eliminate 2, 4, 6, and 8, since any bumps resulting from a first move completely disambiguates. For instance, $\bar{\mathrm{U}}$ is sufficient to show that I am in 2. Since only 5 remains, I have discovered my position. The absurdity of the supposition is made further manifest by the existence of eight other arguments with eight distinct conclusions as to my initial position. For example, I could conclude that I am in 6 by first

> eliminating the corners, then 2, 4, 8, and then 5 (by the sequence L, $\bar{L}$, leaving only 6 remaining). If one individuates arguments by distinct orders of elimination, there are indeed more than eight arguments. I could also conclude that I am in 6 by eliminating in this order: 7, 4 (by U, $\bar{L}$), 8, 1, 2, 5, 9, and 3. Thus I cannot be put in an undiscoverable position.

The sacrificial virgin paradox is intended to show that the subjects need not know how many alternatives there are.

Every fifty years the inhabitants of a tropical paradise sacrifice a virgin to the local volcano in an elaborate ceremony. Virgins from all around are blindfolded and brought before the volcano. They all hold hands in a line and can only communicate one sentence: 'No one to your right is the sacrificial virgin'. This sentence can only be signalled by squeezing the hand of the virgin to one's left. The virgins are reliable and dutibound to so signal if and only if it is known to be true. Besides all this, the virgins also know that a necessary condition for being the sacrificial virgin is that one remain ignorant of the honour until one is tossed in. The chief must take the leftmost virgin up to the mouth of the volcano, and if the offering is acceptable, push her in and tell the rest of the virgins to go home. If the offering is unacceptable, he sends that virgin home and repeats the procedure with the new leftmost virgin. This procedure continues until one virgin is sacrificed, so it is known that one will be sacrificed. After hearing the announcement that one virgin will be sacrificed, someone objects that the ceremony cannot take place:

> The rightmost virgin knows she is rightmost since her righthand is free. She knows that if she is offered, then none of the virgins to her left have been sacrificed. So if she is the sacrificial virgin then she will have to be offered knowing that she is the only alternative remaining, and thus would know she is the sacrificial virgin. Since the sacrificial virgin must not know, the rightmost virgin knows that she is not the sacrificial virgin. This knowledge obliges her to squeeze the hand of the virgin to her immediate left signaling the sentence 'None of the virgins to your right is the sacrificial virgin'. This virgin is either the leftmost virgin or a middle virgin (a middle virgin is any virgin between the leftmost and rightmost virgins). If she is a middle virgin, she will reason that if she is offered, she will know that none of the virgins to her left have been sacrificed. By the signal she knows that none to her right are sacrificial virgins, and thus she will be able to deduce that she will be sacrificed. But since the sacrificial virgin cannot know she will be sacrificed, this middle virgin knows she will not be sacrificed. Therefore, she will squeeze the hand of the virgin to her left, triggering the same deduction if this third virgin is a middle virgin. Once the leftmost virgin is reached, she will know that none of the virgins to her right is the sacrificial virgin since she is the only alternative left. However, she would then both know and not know she is the sacrificial virgin. Therefore, the ceremony is impossible.

The rightmost and leftmost virgins only know that there are at least two virgins in line. Middle virgins only know that there are at least three virgins in line. In the versions previously considered it is essential that the subjects know what the alternatives are. In the surprise examination version the students need to know that examination is on one of the five week days. In the sacrificial virgin version the subjects do not even have a rough estimate as to how many alternatives there are. Middle virgins only know that they are somewhere in the middle of an arbitrarily long line. So it is not essential that the subjects know the order in which members of the series are arranged. Unlike the designated student paradox, the middle virgins *repeat* the same deduction but do not *replicate* each other's deductions. No middle virgin knows more than any other middle virgin. Unlike the other versions, there is no characteristic deduction for each subject. In the designated student paradox, Don can only eliminate himself by replicating Eric's reasoning: Carl can only eliminate himself by replicating Don's replication of Eric's reasoning, and so on. As in the surprise examination paradox, each virgin replicates the reasoning of her 'future self' but not the reasoning of others.

To summarise, these three variations of the prediction paradox show that the temporal retention principle, the KK principle, the order of elimination, and knowledge of the number of alternatives to be eliminated, are each inessential to the prediction paradox. Since past analyses of this paradox do assume the above are essential, I conclude that we do not yet have a formulation of the structure of the prediction paradox.

Michigan State University, East Lansing Received September 1981