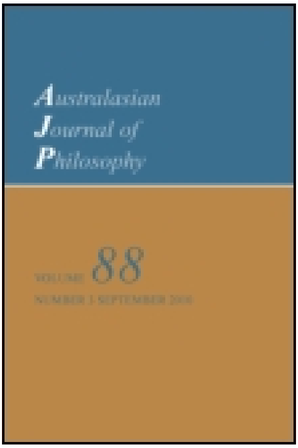

Australasian Journal of Philosophy

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

Conditional blindspots and the knowledge squeeze: A solution to the prediction paradox

Roy A. Sorensen [a]

[a] Michigan State University

Available online: 02 Jun 2006

To cite this article: Roy A. Sorensen (1984): Conditional blindspots and the knowledge squeeze: A solution to the prediction paradox, Australasian Journal of Philosophy, 62:2, 126-135

To link to this article: http://dx.doi.org/10.1080/00048408412341321

Australasian Journal of Philosophy
Vol. 62, No. 2; June 1984

CONDITIONAL BLINDSPOTS AND THE KNOWLEDGE SQUEEZE: A SOLUTION TO THE PREDICTION PARADOX*

Roy A. Sorensen

In 'Recalcitrant Variations of the Prediction Paradox',[1] I introduced three variations of the prediction paradox designed to show that the prediction paradox is still unsolved and that its structure has not yet been formulated. My hope is that this paper will change this state of affairs. In the first section, I present a popular version of the paradox and consider one of the two most influential approaches to the prediction paradox: defending the thesis that it is a variation of the liar paradox. Reasons are then provided for doubting that such a defence could succeed. Section 2 contains a description of the second most influential approach to the prediction paradox. Although faults are found with the past proposals in this tradition, I place my own proposal within it. The third section contains a definition of epistemic blindspots and an explanation of their importance. In the final section, conditional blindspots are defined and their role in the prediction paradox is shown.

1. *The Prediction Paradox and the Liar Paradox*

The most popular version of the prediction paradox involves a teacher who tells his students that there will be a surprise examination next week. A clever student objects that the test is impossible. He first notes that the test cannot be given Friday since the students would then know on Thursday evening that the test must be Friday. The test cannot be given on Thursday because the students would then know on Wednesday evening that the test is either Thursday or Friday, and they have already eliminated Friday. In a like manner, the remaining days of the week are eliminated thereby 'proving' that the test cannot be given.

The possibility of treating the prediction paradox as a variation of the liar paradox began to emerge with Scriven's insistence that the unexpectedness of the test be given a logical rather than a psychological interpretation.[2] The logical interpretation Scriven has in mind is that the test is unexpected in the sense that the students cannot produce a proof that it will occur on a given day. This interpretation leads to the first attempt to show that the prediction paradox is a variation of the liar in R. Shaw's 'The Paradox of the

* This paper has benefitted from the comments of an anonymous referee. Parts of this paper have been taken from my dissertation 'Moore's Problem and the Prediction Paradox'.

[1] Roy A. Sorensen 'Recalcitrant Variations of the Prediction Paradox' *Australasian Journal of Philosophy* (December, 1982). I have not repeated the three variations in this paper since the gist of my solution can be conveyed without them.

[2] Michael Scriven 'Paradoxical Announcements' *Mind* (July 1951).

Unexpected Examination'. Shaw insists that "'knowing' that the examination will take place on the morrow" must be 'knowing' in the sense of 'being able to predict, *provided* the rules of the school are not broken'.[3] Given that 'unexpected' means 'not deducible from certain specified rules of the school', Shaw believes he can formulate two rules for a school having a mandatory surprise examination some day in the term.

> Rule 1: An examination will take place on one day of next term.
> Rule 2: The examination will be unexpected, in the sense that it will take place on such a day that on the previous evening it will not be possible for the pupils to deduce *from Rule 1* that the examination will take place on the morrow.[4]

Although a last day examination can be eliminated since it would violate Rule 2, an examination on any other day would satisfy Rules 1 and 2. By adding a third rule, the possibility of an examination on the last two days can be eliminated.

> Rule 3: The examination will take place on such a day that on the previous evening it will not be possible for the pupils to deduce *from Rules 1 and 2* that the examination will take place on the morrow.[5]

If only two days remain in the term, the pupils can deduce by Rule 1 that the examination is on one of the two remaining days. By Rule 2, they can eliminate the last day, leaving the next to the last day as the only possibility. Since this deduction would violate Rule 3, the last two days are not possible examination days. However, an examination on any other day of the term would satisfy Rules 1, 2, and 3. In general, the last n days of the term are eliminated by appealing to Rule 1 and n additional rules of the form

> Rule n + 1: The examination will take place on such a day that on the previous evening it will not be possible for the pupils to deduce from the conjunction of rules 1, 2, . . ., n, that the examination will take place on the morrow.

The n + 1 rules are incompatible with an n + 1 day term.

Shaw concludes that the original paradox arose by taking in addition to Rule 1,

> Rule 2*: The examination will take place on such a day that on the previous evening the pupils will not be able to deduce from *Rules 1 and 2** that the examination will take place on the morrow.[6]

By applying rule 1 and 2*, one can eliminate every day of the term. Once we realise that 2* is self-referential, the paradox is resolved.

Ardon Lyon complains that Shaw's choice of the rules for the school is

[3] R. Shaw 'The Paradox of the Unexpected Examination' *Mind* (July 1958) p. 386.
[4] *Ibid.* p. 383.
[5] *Ibid.*
[6] *Ibid.* p. 384.

an evasion rather than a solution of the paradox.[7] Lyon points out that mere self-referentiality is not sufficient for paradox. For example, 'This sentence is written in black ink' is perfectly unparadoxical.

A more modest, more technical, attempt to show that the prediction paradox is a variation of the liar appears in David Kaplan's and Richard Montague's 'A Paradox Regained'. They are concerned with another popular version of the prediction paradox; the Hangman. A man, K, is sentenced to hang on one of the following seven noons but must be kept in ignorance until the morning before the execution. The man argues that he cannot be hung on the last day since he would know after the penultimate noon. Having eliminated the last day, the rest are eliminated in the familiar way. According to Kaplan and Montague, the self referential aspect of the judge's decree is reflected by the following reformulation:

> *Unless K knows on Sunday afternoon that the present decree is false,* one of the following conditions will be fulfilled: (1) 'K is hanged on Monday noon' is true, or (2) K is hanged on Tuesday noon but not on Monday noon, and on Monday afternoon K does not know on the basis of the present decree that 'K is hanged on Tuesday noon' is true.[8]

Kaplan and Montague are able to show that this version is a complicated variation of the Liar paradox leading to the conclusion that the decree can and cannot be fulfilled. They go on to consider a one-day version of this variation:

> Unless K knows on Sunday afternoon that the present decree is false, the following condition will be fulfilled: K will be hanged on Monday noon, but on Sunday afternoon he will not know on the basis of the present decree that he will be hanged on Monday afternoon.[9]

Finally, they consider a version in which 'the number of possible dates can be reduced to zero'. Here the judge asserts:

> K knows on Sunday afternoon that the present decree is false.[10]

Unlike Shaw, Kaplan and Montague do not claim that one *solves* the prediction paradox by showing it has an element of self-reference. They merely maintain that by showing that the prediction paradox is a variation of the liar paradox, one has reduced two mysteries to one. Although it is preferable to have a solution, a paradox reduction would still be valuable. For if a solution to the liar paradox was later discovered, one would thereby have a solution to the prediction paradox as well. In any case, the reduction would show that those who seek a solution to the prediction paradox should first seek a solution to the liar.

But despite its popularity amongst formalists, the attempted reduction is beset by a number of difficult objections. First, there is Nerlich's charge that

[7] Ardon Lyon 'The Prediction Paradox' *Mind* (October 1959).
[8] David Kaplan and Richard Montague 'A Paradox Regained' *Notre Dame Journal of Formal Logic* (July 1960) p. 84.
[9] *Ibid.* p. 87.
[10] *Ibid.*

self-reference is not an essential feature of the prediction paradox. Nerlich points out that Shaw provided a non-self-referential formulation of the school rules and a self-referential formulation. Shaw then argued that the first formulation is not paradoxical since no unexpected examination can be given during the term and that the second formulation is paradoxical. Nerlich insists that both formulations are paradoxical. After all, if an examination is given Wednesday, it would not be expected. Thus Shaw's first formulation shows that self-reference is not an essential feature of the prediction paradox.

Second, there is the complaint that there is no self-referential element in the prediction paradox announcements. If there were, then any statement addressed to you describing your ignorance would be self-referential. (B) does not follow (A).

(A) You do not know the capital of Texas.
(B) You do not know the capital of Texas even if you know (B).

It may be perverse of me to say to you

(C) Austin is the capital of Texas but you do not know it,

but (C) is certainly not a contradiction. In fact, (C) would be true if my utterance of (C) left you in ignorance of Texas' state capital. But if we follow the self-referentialists in equating knowability with deducibility, (C) would be a contradiction. For (C) would imply

(C') Austin is the capital of Texas but this fact cannot be deduced from this sentence.

The self-referentialist does not do justice to the intuition that the announcement is satisfiable and satisfiable after the teacher has announced it. Nor does the self-referentialist do justice to the intuition that the students are informed by the announcement. These objections add fuel to the suspicion that the self-referentialists have attempted to treat the problem by distorting its form.

2. *Past Attacks on the Base Step*

I think the paradox can be solved by showing that the students can be surprised by a Friday test. Showing the soundness of this objection to such a persuasive step in the argument is a formidable task. Quine attacked this step on the grounds that the students can still be surprised because they do not really know that the teacher's announcement is true.[11] This proposal is unattractive because the refusal to concede that the students know the announcement is true seems to require a scepticism about knowledge by authority. Intuitively, the students are informed by the announcement. The overthrow of this intuition requires a powerful argument.

In 'The Suprise Examination in Modal Logic',[12] Robert Binkley seems

[11] W. V. Quine 'On a so-called Paradox' *Mind* (January 1953).
[12] Robert Binkley 'The Surprise Examination in Modal Logic' *The Journal of Philosophy* (March 7, 1968).

to provide the required argument. Binkley points out that the announcement corresponding to the 1 day version of the paradox.

(1) There will be a test tomorrow but you will not know the day in advance,

resembles the sentence G. E. Moore was so perplexed by:

(2) It is raining but I don't believe it.

In *Knowledge and Belief,* Jaakko Hintikka argued that (2) cannot be believed by a perfect logician even though it is consistent. Since the prediction paradox is a paradox involving perfect logicians, Binkley points out that Hintikka's explanation of the incredibility of (2) can be extended to the question of why the students cannot know (1). The students cannot believe, and therefore, cannot know (1) because (1) is logically incredible to the students. By appealing to the principle that if a perfect logician believes p, then he believes that he will believe p thereafter, Binkley is able to demonstrate that any n + 1 day announcement must also be incredible to the students. So Binkley concludes that the prediction paradox is in the same family as Moore's problem.

In 'The Paradox of the Unexpected Examination'[13] Crispin Wright and Aidan Sudbury agree with Quine and Binkley that the students cannot know the announcement corresponding to the 1 day case. They disagree on the question of whether the students can know the announcement corresponding to the n + 1 day case. According to Wright and Sudbury, the students can reasonably believe the n + 1 day announcement as long as there are n days left. After all, there is nothing wrong with believing on Sunday that on one of the next five days it will be the case that 'Today is the examination day but I did not believe so last night'. However, if the test is not given by Thursday, there is something wrong in believing that on Friday it will be true that 'Today is the examination day but I did not believe so last night'. The problem is Moore's problem since one would in effect be believing that it is both the case that there will be an examination tomorrow and that one does not believe it. Thus the teacher's announcement makes a hiatus in reasonable belief possible. People who are not surprisees are not vulnerable to this hiatus since Moorean sentences implied by the fact that the examination will be a surprise are not about them. According to Wright and Sudbury, the teacher's announcement can therefore be fulfilled while doing justice to the intuitive meaning of the announcement. Their proposal allows the students to be informed by the announcement and explains why the paradox does not arise if the teacher keeps the announcement to himself or only tells someone other than the students.

The essence of the Wright/Sudbury solution can be stated simply: reject the temporal retention principle. The temporal retention principle states that if one knows or reasonably believes p, one will do so thereafter. This principle has been employed by commentators who are aware of the fact that the principle does not hold in the everyday world where people forget, die and go

[13] Crispin Wright and Aidan Sudbury 'The Paradox of the Unexpected Examination' *Australasian Journal of Philosophy* (May, 1977).

insane. They defend their use of it on the grounds that these events are irrelevant contingencies which can be stipulated away to reveal a paradox for ideal thinkers in favourable epistemological environments. Wright and Sudbury seem sympathetic to such idealisations. They base their rejection of the temporal retention principle on the fact that many Moorean sentences are incredible only at certain times. On Monday there is nothing wrong with me believing that 'On Tuesday, it will be raining but I won't believe it', but on Tuesday the sentence is logically incredible to me. By bringing out the Moorean nature of the teacher's announcement. Wright and Sudbury provide an explanation of why the temporal retention principle must be rejected.

As asserted in 'Recalcitrant Variations of the Prediction Paradox', I agree with the rejection of the temporal retention principle but I do not think that one thereby solves the prediction paradox. There is a variation, the designated student paradox, that does not require appeal to the temporal retention principle. Since an adequate solution to the prediction paradox must be complete in the sense that it must be a solution for all of the variations, the proposal advanced by Wright and Sudbury is not an adequate solution.

3. *Epistemic Blindspots*

Crucial to my solution of the prediction paradox is the concept of an epistemic blindspot.[14] A proposition p is an *epistemic blindspot* for person a (at time t) if and only if p is consistent, while Kap (for a knows that p) is inconsistent. For example, 'It is raining but Bob does not know so' is an epistemic blindspot for Bob but not for Cal. This can be shown by letting 'q' stand for 'It is raining', 'b' denote Bob, and 'c' denote Cal. Although 'It is raining but Bob does not know so' is consistent, the supposition that Bob knows that 'It is raining but Bob does not know so' leads to an inconsistency.

1. $Kb(q \;\&\; \text{-}Kbq)$	Assumption.
2. $Kbq \;\&\; Kb\text{-}Kbq$	1, Knowledge distributes over conjunction.
3. $Kbq \;\&\; \text{-}Kbq$	2, TF, Knowledge implies truth.

However, no contradiction follows from the supposition that $Kc(q \;\&\; \text{-}Kbq)$. Thus 'It is raining but Bob does not know so' is an epistemic blindspot for Bob but not for Cal.

One reason why epistemic blindspots are important is that they are counterexamples to certain generalisations about the scope of knowledge. Among these generalisations the following are of interest.

(UA) Unrestricted access principle: Whatever can be true, can be known; $(p)(\Diamond p \supset (x)\Diamond Kxp)$.

(IT) Intertemporal access principle: Whatever can be known to a person at one time, can be known to him at any other time; $(p)(x)(t_1)(t_2)(\Diamond Kxt_1p \supset \Diamond Kxt_2p)$.

[14] My concept of an epistemic blindspot was derived from Jaakko Hintikka's concept of an anti-performatory statement described in his *Knowledge and Belief* (Ithaca: Cornell University Press, 1962) pp. 90-91.

(IP) Interpersonal access principle: Whatever can be known by someone, can be know by anyone else; $(p)(x)(y)(\diamond Kxp \supset \diamond Kyp)$.

Although (UA) is tempting, it conflicts with a major part of the epistemological enterprise; describing the limits of what we can know. For example, Kant claimed that propositions about noumena are unknowable and Moses Maimonides claimed that positive propositions about God are unknowable. Despite the controversiality of the limits asserted by Kant and Maimonides, one can see that they were correct in their rejection of (UA) by considering the universal epistemic blindspot:

(3) No one knows anything.

Since $-(\exists y)(\exists p)Kyp$ is consistent and $(\exists x)Kx\text{-}(\exists y)(\exists p)Kyp$ is inconsistent, (3) is an epistemic blindspot for everyone.

Although the intertemporal access principle reflects the intuition that time is epistemically irrelevant, its falsity can be shown with the following humdrum blindspot.

(4) John will first know that he was adopted on his eighteenth birthday.

Contrary to (IT), John can know (4) when he is nineteen but cannot know it when he is seventeen. This counterexample provides support for the rejection of the temporal retention principle. Sometimes one cannot be in a position to know simply because of one's temporal position.

Just as when you are seems irrelevant to what you can know, who you are seems epistemically irrelevant. Despite this intuition, there are counterexamples to the interpersonal access principle.

(5) John never did, does not, and never will know that he was adopted.

Contrary to (IP), John's mother can know (5) but John cannot. This second observation also supports the analysis given by Wright and Sudbury. One counterintuitive consequence of their proposal is that nonstudents can know that the test will be given on the only remaining day, but the students cannot. Despite the fact that the students and nonstudents have the same evidence and are ideal thinkers, the nonstudents as nonsurprisees are in a position to know something that the students cannot. This strange result is due to the fact that the conjunction of the teacher's announcement and the proposition that only one day remains is an epistemic blindspot for the students but not for nonstudents (since the announcement only says that the *students* will be surprised).

4. *Conditional Blindspots*

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. For example,

(6) If Ralph survived, he is the only one who knows it.

is a conditional blindspot for everyone except Ralph.

Given that a proposition is a conditional blindspot to you, it is possible for you to know the proposition and it is possible for you to know its antecedent. However, it is impossible to know both the conditional blindspot and its antecedent. To see this, recall the proof for the unknowability of epistemic blindspots and let '*r*' be 'Bob is drugged'.

1. $Kbr \,\&\, Kb(r \supset (q \,\&\, \text{-}Kbq))$	Assumption.
2. $Kbr \supset Kb(q \,\&\, \text{-}Kbq)$	1, TF, Knowledge distributes over conditionals
3. $Kb(q \,\&\, \text{-}Kbq)$	1,2, TF

Since it has already been shown that a contradiction can be derived from the supposition that $Kb(q \,\&\, \text{-}Kbq)$, we can see the inconsistency of 'Bob knows he is drugged and knows that if he is drugged, it is raining but he does not know it'.

Once one realises that the teacher's announcement for the n + 1 day case is a conditional blindspot for the students (but not nonstudents), an adequate solution to the prediction paradox begins to emerge. For the sake of simplicity, consider the 2 day case of the teacher's announcement involving only one student, Dave.

(7) Either the test will be given Thursday or Friday but in neither case will Dave know in advance.

By letting 'p_4' stand for 'The test is given Thursday' and letting 'p_5' stand for 'The test is given Friday' while '*d*' denotes Dave, the announcement can be symbolised as

(8) $(p_4 \,\&\, \text{-}Kdp_4) \vee (p_5 \,\&\, \text{-}Kdp_5)$.

To see that (7) is a conditional blindspot for Dave, note that (8) is equivalent to

(9) $\text{-}(p_4 \,\&\, \text{-}Kdp_4) \supset (p_5 \,\&\, \text{-}Kdp_5)$.

Since the announcement is only a conditional blindspot for Dave, nothing prevents him from knowing it on the basis of the teacher's authority. However, once Dave learns that no test has been given Thursday, he knows the antecedent of (9) and so cannot continue to know the announcement. Dave's friend Fred, could continue to know (9) and the antecedent of (9) since it is not a conditional blindspot to Fred. Thus the teacher can inform Dave with announcement (7) and give the test on Friday.

One might object that in the event only one day remains, Dave would be more likely to believe that a test (albeit unsurprising) will be given on the remaining day than not at all. This may be the case since, empirically, most (but not all) teachers would prefer to make part of the announcement true than give no test at all. However, Dave does not have this extra psychological information about his teacher's preferences. It would be pointless to grant Dave this extra information since it would only serve to undermine the clever student's argument against the possibility of a Friday test.

The conditional blindspot analysis can be extended to all variations of the prediction paradox. All n + 1 step variations involve a knowledge squeeze, that is, a situation in which one's knowledge of a conditional blindspot comes in conflict with one's knowledge of its antecedent. The general pattern of a prediction paradox argument runs as follows. First, one describes a situation in which a person or group, *a*, appears to know a proposition which is a conditional blindspot for *a*. In the above case, this is done by describing how Dave's trustworthy teacher tells him that the conditional blindspot is true. Second, one describes a contingency in which *a* appears to learn that the antecedent of the conditional blindspot is true. This is exemplified by Dave observing that no test has been given on Thursday. Third, one asks 'Is this contingency really possible?'. If this question is interpreted as 'Is it really possible for *a* to know both his conditional blindspot and its antecedent?', the correct answer is 'No'. If the question is interpreted as 'Is it really possible for both the conditional blindspot and its antecedent to true?', the correct answer is 'Yes'. But due to unfamiliarity with the peculiar epistemic behaviour of conditional blindspots, victims of the paradox confidently concede that the contingency really is not possible and accept the consequences that follow from a negative answer to the second reading of the question. The consequence is that the contingency is eliminated and the negation of the consequent is added to our background knowledge. Thus, we agree that Dave can eliminate the possibility of a Friday test, and allow Dave to use this knowledge in his examination of other contingencies. Fourth, one shows how this derived knowledge of the negation of the consequent puts *a* in a position to eliminate another contingency. This domino effect results from the fact that the conjunction of the conditional blindspot with the negation of its consequent is either a blindspot or a new conditional blindspot. If the conjunction is a blindspot, one asks 'Is this really a contingency?'. Once again, the question has two readings: 'Is it possible for *a* to know that his blindspot is true?' and 'Is it possible for the blindspot to be true?'. And once again, unfamiliarity with epistemic blindspots leads the victim of the paradox to incorrectly rule out this last possibility and conclude that the announcement cannot be fulfilled. This is what happens in the case of Dave. After eliminating the possibility of a test on Friday, he is faced with the blindspot that the test is on Thursday but he does not know the test is on Thursday. The victim of the paradox reasons that since one cannot surprise someone as to when something will happen if one tells him when it will happen, Dave cannot be surprised by a Thursday test. On the other hand, if the conjunction of the conditional blindspot and the negation of its consequent is a new conditional blindspot then the process of deriving the negation of the consequent of a conditional blindspot is repeated on this new conditional blindspot.

As illustrated by the recalcitrant variations of the prediction paradox, the knowledge squeeze can be set up in a variety of ways by varying the number and kind of epistemic opportunities and by varying the nature of the blindspots contained in the conditional blindspots. In the paradox of the undiscoverable position, there are four possible base steps and a large number of ways of reducing the initial conditional blindspot to a blindspot. So here, unlike the traditional variations of the prediction paradox, there is no rigid order of elimination. In the designated student paradox, the initial conditional blindspot contains individual blindspots for each of the students. The order of elimination is dictated by the increasing perceptual knowledge enjoyed by students at the rear of the line. In the sacrificial virgin paradox, conditional blindspots come into play as universal instantiations of the general description of how a ceremony will take place. Epistemic opportunities to eliminate contingencies arise from a signalling system and the ability to feel whether one has a neighbouring virgin.

Despite this variety, there is a single feature shared by all of the prediction paradoxes that provides the basis for a general solution. All variations of the paradox turn on fallacious reasoning about blindspots or conditional blindspots. In the 1 step cases, for example the 1 day case of the surprise examination paradox, one becomes a victim of the paradox only if one falsely assumes that blindspots can be known by their holders. In the n + 1 step cases, conditional blindspots are used to set up a knowledge squeeze. Here, one becomes a victim only if one fallaciously infers the negation of a conditional blindspot's consequent from the impossibility of knowing both the conditional blindspot and its antecedent.

Michigan State University Received May 1983