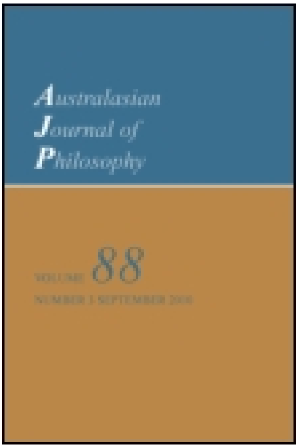

Australasian Journal of Philosophy

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

The paradox of the unexpected examination

Crispin Wright [a] & Aidan Sudbury [b]

[a] All Souls' College, Oxford

[b] Monash University

Available online: 18 Sep 2006

Australasian Journal of Philosophy
Vol. 55, No. 1; May 1977

CRISPIN WRIGHT AND AIDAN SUDBURY

THE PARADOX OF THE UNEXPECTED EXAMINATION

I

We had better begin by rehearsing the puzzle. During a Sunday evening assembly at a boarding school a headmaster announces to the fourth form that at 9 a.m. on one of the next six days they will be set a surprise examination, a surprise in the sense that on the evening before the examination actually occurs it will not be possible to predict its occurrence on the following day. The pupils argue, ingeniously, that the headmaster cannot carry out the announcement since the conditions he has imposed are incoherent. They reason like this:

1. If the examination were to be held over till Saturday, the last available day, then on Friday evening we should be in a position to predict its occurrence on the following day.

2. However we will not, apparently, ever be in such a position.

Hence:

3. The examination will occur before Saturday.

4. However, since Saturday is excluded, it is clear that an examination on Friday will not comply with the announcement either. For having excluded Saturday we should be in a position on Thursday evening to predict a Friday examination, contrary to the announcement.

Hence:

5. The examination will occur before Friday . . . etc.

The pupils thus proceed to eliminate all the available days, concluding that one part of the headmaster's announcement—that there will be an examination in the week in question—is inconsistent with the other—that they will not be in a position to predict its date on the evening before it occurs. The reasoning has some plausibility, but the conclusion seems manifestly unsound. If the headmaster decides, *e.g.* to set the examination on Wednesday, it seems clear that the pupils need have no reason to expect as much on Tuesday evening. So we have a paradox in the classic sense: apparently sound reasoning has led to an apparently unacceptable conclusion; specifically, an arguably inconsistent set of propositions appears to be satisfiable.

It would be intuitively satisfying to have a solution to the puzzle which met the following six conditions:

(A) The account given of the content of the announcement should make it clear that it *is* satisfiable, since a surprise examination is, palpably, a logical possibility.

(B) The account should make it clear that the headmaster can carry out the announcement even *after* he has announced it since, palpably, he can.

These two conditions require that the paradox not be construed as straightforwardly one of impredicativity or 'pragmatic self-refutation'.

(C) The account must do justice to the intuitive meaning of the announcement. An extraordinary proportion of commentators have chosen to discuss quite unnatural interpretations of it.

(D) The account must do justice to the intuitive plausibility of the pupils' reasoning.

(E) The account should make it possible for the pupils to be *informed* by the announcement: we want the reaction of someone who notices no peculiarity but just gets on with his revision to be logically unobjectionable.

(F) The account must explain the role, in the generation of the puzzle, of the announcement's being made to the *pupils*; there is, intuitively, no difficulty if, *e.g.* the headmaster tells only the second master or keeps his intentions to himself. Most of the interpretations in the literature which identify the problem as one of impredicativity fail to meet this condition.

None of the discussions with which we are familiar meets all these conditions. That, of course, is no objection unless they can all be met. The purpose of this paper is to suggest how this may be possible.

II

We need a more formal treatment of the pupils' reasoning, for it seethes with inexplicit assumptions. In particular, as Quine[1] noted, it is bogus as a straightforward *reductio* of the actual announcement, for the manner in which Saturday is eliminated presupposes that the pupils would know or have reason to believe *on Friday* that an examination was still going to occur. Thus the pupils assume, beyond what is announced, that they know or have reason to believe the announcement and will continue to do so. Quine thought the pupils' reasoning demonstrated the inconsistency of the announcement with their knowledge of it; if that is correct, there is no hope for condition E. Ayer,[2] in any case, invented a variant of the

[1] 'On a so-called paradox', *Mind* 62 (1953); reprinted in *The Ways of Paradox* (1966).

[2] 'On a supposed Antimony', *Mind* 82 (1973).

paradox which virtually eliminates doubts as to the eventual occurrence of the 'surprise' event. In his example someone is told that when a pack of cards is dealt out, he will not know beforehand when the Ace of Spades will arrive; but he is allowed to check for its presence before the pack is shuffled. What is clear for our immediate purpose, however, is that we shall need some sort of epistemic operator in order to represent the pupils' reasoning formally. We shall write this as: $D^x_t p$, usually to be paraphrased as: 'Proposition(s) p are at the disposal of person(s) x at time t'. The point of this 'disposal' terminology is merely to avoid an early judgement about what epistemic notion(s) may be involved.

Further, 'E_m' will express that the examination occurs on day m, and 'A_m' the exclusive disjunction of $\{E_1, \ldots, E_m\}$. Thus A_n is part of the content of the headmaster's announcement, where n is the number of days in the week.

The pupils' reasoning implicitly attributes a variety of properties to the epistemic notion involved. They assume that what they have been told is at their disposal, that consequences of propositions at their disposal are likewise at their disposal, that propositions which they subsequently verify will be at their disposal and, as just noted, that certain propositions at their disposal will continue to be so. Let us codify these presuppositions for '$D^x_t p$':

d(i): if x is informed of p at t, then $D^x_t p$.

d(ii): if a set of propositions, $\{q_1, \ldots, q_m\}$, entails p and for each q_i in that set, $D^x_t q_i$, then $D^x_t p$.

In what is to come '$D^x_t\{q_1, \ldots, q_m\}$' is to be taken, in the obvious way, as true if and only if $D^x_t q_i$ is true for each q_i in $\{q_1, \ldots, q_m\}$. We shall cite appeals to this as appeals to *d*(ii).

d(iii): if the experience of x at t constitutes verification of p, then $D^x_t p$. Thus in particular if the examination does not occur on or before day m, then $D^s_m \sim A_m$, where 's' denotes the pupils.

Finally, let us take it that if a proposition ever comes to be at someone's disposal, then it remains so:

d(iv): for any r, t, $r < t$, if $D^x_r p$, then $D^x_t p$.

We are not yet concerned with how plausible these principles are under any intuitive epistemic interpretation. They are rather to be regarded as *analytic* of the notion of having a proposition at one's disposal, and accordingly will feature as rules of inference in the ensuing interpretation of the pupils' reasoning as a piece of natural deduction. Note one conscpicuous absence: we do not have, nor shall we require, that '$D^x_t p$' entails the *truth* of 'p', the distinctive feature of knowledge.

Specifically, we have the following rules for sequent derivation. For *d*(ii):

$$: \quad \frac{\Gamma \vdash \Delta;\ \Theta \vdash D_t^x\Gamma}{\Theta \vdash D_t^x\Delta},$$

and

$$: \quad \frac{\Gamma \vdash D_t^x\{q_1, \ldots, q_m\}}{\Gamma \vdash D_t^x q_1, \ldots, D_t^x q_m}.$$

(Under this heading, we shall make special use of the following pair of derived rules:

$$\frac{\Gamma \vdash \Theta;\ \Delta \vdash \sim D_t^x\Theta}{\Delta \vdash \sim D_t^x\Gamma},$$

and:

$$\frac{\Gamma \vdash \sim D_t^x\{q_1, \ldots, q_m\};\ \Delta \vdash D_t^x\{q_1, \ldots, q_i\}}{\Gamma, \Delta \vdash \sim D_t^x\{q_i, \ldots, q_m\}}.)$$

For d(iii):

$$\frac{\Gamma \vdash \sim A_m}{\Gamma \vdash D_m^s \sim A_m}.$$

For d(iv):

$$\frac{\Gamma \vdash D_r^x\Delta}{\Gamma \vdash D_t^x\Delta},\ r < t.$$

For reasons which will emerge we formulate no rule for d(i).

How, then, in these terms should we formulate the notion of surprise? Our concern is with what the pupils, roughly, have a right or are in a position to expect. Intuitively, an event comes as a surprise in this sense if at no time before it occurs do we have available information of sufficient logical strength to predict the time of its occurrence. An examination on day m will thus surprise the pupils if and only if no set of propositions, Γ, which entails the true 'E_m', will be at their disposal before m. So the following expression of residual content of the headmaster's announcement suggest itself:

$$C\text{:} \quad (m)(\Gamma)[(\Gamma \text{ entails } E_m) \rightarrow (\sim E_m \vee \sim D_{m-1}^s\Gamma)].$$

In the presence of the rule, d(ii), this evidently reduces to:

$$C^*\text{:} \quad (m)[E_m \rightarrow \sim D_{m-1}^s E_m].$$

This is the version which we shall use.

Let us now consider how to present the pupils' reasoning as a natural deduction. We will utilise any standard underlying truth-functional and quantificational logic, and will signal explicitly only applications of the d-rules. We assume the truth of the announcement:

1	(1) A_n	Assumption.
2	(2) C^*	Assumption.

and that these propositions have been announced to the pupils, whence by d(i):

3	(3) $D^s_0 A_n$	
4	(4) $D^s_0 C^*$	

For convenience of later discussion, we take 3 and 4 as separate assumptions. We now proceed:

5	(5) $\sim A_{n-1}$	Assumption
1, 5	(6) E_n	(1), (5).
1, 2, 5	(7) $\sim D^s_{n-1} E_n$	(2), (6).
1, 2, 5	(8) $\sim D^s_{n-1}\{A_n, \sim A_{n-1}\}$	(6), (7), d(ii).
3	(9) $D^s_{n-1} A_n$	(3), d(iv).
1, 2, 3, 5	(10) $\sim D^s_{n-1} \sim A_{n-1}$	(8), (9), d(ii).
1, 2, 3, 5	(11) A_{n-1}	(10), d(iii).
1, 2, 3	(12) A_{n-1}	(5), (11).

This completes the first stage of the reasoning. Notice that no use has so far been made of Assumption 4. We have ruled out the last day, as is intuitively proper, just on the assumptions 1 (that the event will take place in the stated period) and 3 (that the pupils have this fact at their disposal) and 2 (that the date will be unpredictable by them). Can we now proceed uniformly to rule out each day, given the additional assumption, 4, that the pupils have also been informed at the outset of C^*?

13	(13) $\sim A_{n-2}$	Assumption.
1, 2, 3, 13	(14) E_{n-1}	(12), (13).
1, 2, 3, 13	(15) $\sim D^s_{n-2} E_{n-1}$	(2), (14).
1, 2, 3, 13	(16) $\sim D^s_{n-2}\{A_n, C^*, D^s_0 A_n, \sim A_{n-2}\}$	(14), (15), d(ii).
3	(17) $D^s_{n-2} A_n$	(3), d(iv).
4	(18) $D^s_{n-2} C^*$	(4), d(iv).

Now, however, we reach an impasse. No way is apparent for proceeding save by means of $D^s_{n-2} D^s_0 A_n$; and no way is apparent for deriving it from initial data. Unfortunately, with natural deduction methods it is not in general possible to prove that a certain assumption is necessary. However, it is intuitively clear that we do not have sufficient premises in order to proceed. Accordingly we take as a further assumption:

19	(19) $D^s_{n-2} D^s_0 A_n$	Assumption
1, 2, 3, 4, 13, 19	(20) $\sim D^s_{n-2} \sim A_{n-2}$	(16, (17), (18), (19), d(ii).
1, 2, 3, 4, 13, 19	(21) A_{n-2}	(20), d(iii).
1, 2, 3, 4, 19	(22) A_{n-2}	(13), (21).

It thus appears that 1, 2, 3, 4 do not, at least within the framework of the stated *d*-rules, suffice even to rule out the penultimate day. And, indeed, further *ad hoc* assumptions are going to be required as the reasoning proceeds:

23	(23) $\sim A_{n-3}$	Assumption.
1, 2, 3, 4, 19, 23	(24) E_{n-2}	(22), (23).
1, 2, 3, 4, 19, 23	(25) $\sim D^s_{n-3}E_{n-2}$	(2), (24).
1, 2, 3, 4, 19, 23	(26) $\sim D^s_{n-3}\{A_n, C^*, D^s_0A_n, D^s_0C^*, D^s_{n-2}D^s_0A_n, \sim A_{n-3}\}$	(24), (25), d(ii)

To proceed to $\sim D^s_{n-3} \sim A_{n-3}$, we now require: $D^s_{n-3}A_n$, $D^s_{n-3}C^*$, $D^s_{n-3}D^s_0A_n$, $D^s_{n-3}D^s_0C^*$ and $D^s_{n-3}D^s_{n-2}D^s_0A_n$. Only the first two are derivable from our initial basis. Thus in addition to the given assumptions, 1–4, the derivation of A_{n-3} is going to require *four* further assumptions: 19 and the latter three just listed. And in general the position is this: to prove A_{n-m}, $m \geqslant 3$, the pupils will need 1–4 *plus* the additional assumptions, Δ, used in proving $A_{n-(m-1)}$, *plus* $D^s_{n-m}\{3,4,\Delta\}$, the latter expression being a shorthand notation for the more precise $D^s_{n-m}(\Delta \cup \{(3)\} \cup \{(4)\})$.

The first lesson of our semi-formalisation, then, is that some rule of iteration for the *D*-operators is going to be required if it is to be possible to carry through the pupils' reasoning just on the basis of 1–4 as assumptions. Actually, the simplest such rule would suffice:

$$\frac{\Gamma \vdash D^x_t\Delta}{\Gamma \vdash D^x_tD^x_t\Delta}.$$

By this principle and d(iv) all the required additional premises, as the reader may verify, can be derived from 3 and 4.

But may such a principle be regarded as analytic? The question cannot be resolved at this stage. We are, after all, in the process of stipulating a sense for the *D*-operators in terms of which it is to be possible to represent the pupils' reasoning. Later it will be in point to ask whether any intuitive epistemic concept can exemplify all the properties assigned to the *D*-operators.

When the pupils' reasoning is presented informally, this is almost always done from their first-person point of view. This is what makes it difficult to see that iterations of this operator will be needed. This difficulty connects with Moore's 'Paradox':

'*p*, but I do not believe that *p*'.

If, by his own admission, someone does not believe that *p*, then presumably he does not consider that he has adequate grounds for believing *p*; and in that case he cannot consider that he is in a position to inform a hearer of the conjunction. The 'paradox' is nearer the surface in:

'*p*, but I am unaware that *p*',
'*p*, but I have no reason to believe *p*',
'*p*, but I am not justified in asserting *p*'.

The truth of these assertions is actually inconsistent with the capacity of the speaker to inform a hearer by means of them; for it is inconsistent with their being made in accordance with what we may (not altogether happily) describe as the convention that a speaker is only to make assertions which he considers himself in a position to justify.

The 'convention' governs not just attempts to inform but assertions made to record findings, to endorse assertions made by others, *etc.* And, corresponding to the paradoxical flavour of Moore's assertion and the others listed, it injects a ring of pleonasm into:

'p, and I am aware that p',
'p, and I am justified in asserting that p',
'p, and I have good reason to believe p'.

In any context in which an assertor is understood to be intending to follow the convention, for him to assert the first conjunct in such examples is to indicate that he considers himself in a position to defend the second; hence there is no point in asserting that as well.

What would appear to happen when we, in the rôle of the pupils, agree with the informal reasoning is that successive conclusions, 'The examination will occur by Friday evening', 'The examination will occur by Thursday evening', *etc.*, are each, when used as premises for the next stage, assigned the logical force of what would standardly be an otiose rider to their assertion, *viz.* 'We are aware, *etc.*, that the examination will occur by Friday evening'. This is easily done because in any ordinary assertoric context nothing is added by the speaker's affirming the rider.

Be that as it may. The situation is, nevertheless, that if q is the proposition with which a particular stage of the pupils' reasoning concludes, 'We are aware that q' is what they will next need; and this will be available—assuming that an analogue of d(ii) captures the mode of transmission of the relevant epistemic prefix across logical consequences—only if 'We are aware' of the truth of the premises which yielded q. So the number of iterations of the prefix on the initial premises will increase by one for each successively eliminated day.

Let us then add to the pool of d-rules:

$$d(\text{v}): \quad \frac{\Gamma \vdash D_t^x \Delta}{\Gamma \vdash D_t^x D_t^x \Delta}.$$

Now the intitial premises of the natural deduction do indeed serve for a potentially paradoxical conclusion. Line (19) may now be justified thus:

3 (19) $D_{n-2}^s D_0^s A_n$ (3), d(iv), d(v).

and (19) accordingly disappears from the list of assumptions at line (22). Thus line (26) becomes:

1, 2, 3, 4, 23 (26) $\sim D_{n-3}^s \{A_n, C^*, D_0^s A_n, D_0^s C^*, \sim A_{n-3}\}$ (24), (25), d(ii).

and we can proceed:

3	(27) $D^s_{n-3}A_n$	(3), d(iv).
4	(28) $D^s_{n-3}C^*$	(4), d(iv).
3	(29) $D^s_{n-3}D^s_0A_n$	(3), d(iv), d(v).
4	(30) $D^s_{n-3}D^s_0C^*$	(4), d(iv), d(v).
1, 2, 3, 4, 23	(31) $\sim D^s_{n-3} \sim A_{n-3}$	(26) to (30), d(ii).
1, 2, 3, 4, 23	(32) A_{n-3}	(31), d(iii).
1, 2, 3, 4	(33) A_{n-3}	(23), (32).

.

.

.

and so on.

How, then, does the reasoning conclude? By continuing as indicated, we shall arrive at:

1, 2, 3, 4	$(k)A_1$	$(k-1)$, $(k-10)$.
1, 2, 3, 4	$(k+1) \sim D^s_0E_1$	(2), (k).
1, 2, 3, 4	$(k+2) \sim D^s_0\{A_n,C^*,D^s_0A_n,D^s_0C^*\}$	(k), $(k+1)$, d(ii).
3	$(k+3)D^s_0D^s_0A_n$	(3), d(v).
4	$(k+4)D^s_0D^s_0C^*$	(4), d(v).
3, 4	$(k+5)D^s_0\{A_n,C^*,D^s_0A_n,D^s_0C^*\}$	(3), (4), $(k+3)$, $(k+4)$, d(ii)
3, 4	$(k+6) \sim \{A_n,C^*\}$	$(k+2)$, $(k+5)$.

We have an inconsistency, then, not *within* the headmaster's announcement, as the pupils originally purported to prove, for there is no longer any reason to think that there is any such inconsistency, but between the announcement and the supposition that it is at the disposal of the pupils. This, of course, is not yet a paradoxical result. We are entitled to no intuitions about whether the headmaster, having made the announcement, can proceed to surprise the pupils relative to all the information *at their disposal*. But we do believe that he can surprise them relative to everything which they know, or have reason to believe. If there is a paradox in the area, it thus depends on whether d(i)-d(v) are analytic of some such intuitive epistemic concept.

III

Let us entertain as an interpretation of 'D^x_tp': x has *good reason to believe p* at t, where this is taken to involve that x's total state of information at t justifies his belief in p and assertion that p. Clearly 'D^x_tp', so interpreted, may be true while p is not, so we are considering something less than knowledge. Further, it is (at least) good reason to believe the announcement, in this weaker-than-knowledge sense, that we want the headmaster to be able to communicate to the pupils.

Perhaps he cannot, and endorsement of condition E is simply a mistake. Here is where it is in point to reflect that we made no appeal to *d*(i) in the deduction but directly assumed $D_0^s A_n$ and $D_0^s C^*$. Suppose *d*(ii) — *d*(v) are valid for reasonable belief. From the sequent with which the deduction concluded:

$$D_0^s A_n, D_0^s C^* \vdash \sim\{A_n, C^*\},$$

we can proceed as follows:

$$D_0^s\{D_0^s A_n, D_0^s C^*\} \vdash D_0^s \sim \{A_n, C^*\}, \quad \text{by } d(\text{ii}).$$
$$D_0^s\{A_n, C^*\} \vdash D_0^s\{D_0^s A_n, D_0^s C^*\}, \quad \text{by } d(\text{v}), d(\text{ii}).$$

Thus: $$D_0^s\{A_n, C^*\} \vdash D_0^s \sim \{A_n, C^*\}.$$

Under the present interpretation, this sequent requires serious thought. It entails that if the pupils are in a position reasonably to believe what the headmaster announces, then they are so placed with respect to its negation also. But one cannot be in a position *reasonably* to believe both *p* and its negation. No situation where the evidence was of differing strengths would warrant such a description; and if the evidence were of the same strength—whatever that might mean—the reasonable thing to do would be to suspend belief in both propositions. Thus for this interpretation of the *D*-operators we may validly avail ourselves of:

$$d(\text{vi}): \quad \frac{\Gamma \vdash D_t^x \Delta;\ \Theta \vdash D_t^x \sim \Delta}{\Gamma \vdash \sim\Theta}.$$

We can now proceed:

$$D_0^s\{A_n, C^*\} \vdash D_0^s\{A_n, C^*\},$$
$$D_0^s\{A_n, C^*\} \vdash \sim D_0^s\{A_n, C^*\}, \quad \text{by } d(\text{vi}).$$

Hence: $$\vdash \sim D_0^s\{A_n, C^*\}.$$

Granted the validity, then, of *d*(ii)-*d*(vi), there is no alternative to concluding that the pupils cannot reasonably believe—*a fortiori,* cannot know—both A_n and C^* at time *0*. Ths is the 'solution' to the puzzle implicit in Quine's discussion.

We can do a little to mitigate its counterintuitiveness. Suppose that *X* reasonably believes *p* and reasonably believes also that *Y* has no good ground to believe *p*; so he attempts to tell him so:

'*p*, but you have no good reason to believe that *p*'.

Evidently *X* does not succeed in giving *Y* good reason to believe what he has said, or *Y* would be in a position reasonably to assert a variant of Moore's paradox. Just for that reason *X* can continue reasonably to believe both conjuncts of the assertion, even after he has made it. Grounds for belief in such an assertion are not capable of communication to *Y*.

Now consider a one-day version of the headmaster's announcement:—

$$1 : A_1$$
$$2 : C^*$$

Obviously, 1 and 2 conjointly entail:—

$$E_1 \ \& \sim D_0^s E_1$$

a version, when directed to s at 0, of the sentence which we have just considered. Thus the headmaster cannot bring about by the announcement, or by any other way, that $D_0^s\{1,2\}$. For that reason his making the announcement on Sunday poses no obstacle to his capacity to carry it out on Monday. Moreover, while the pupils ought to recognize, even after the announcement, that the headmaster can still implement it, they have to acknowledge that his capacity to do so turns on the fact that, lacking any further background information, the logically appropriate reaction on their part to the announcement was one of bewilderment. Thus if all they have to go on is the announcement, they cannot regard themselves as the recipients of a sincere expression of intention.

The suggestion, then, is that the n-day announcement is essentially comparable to the one-day version. It is one of a genre of assertions which express the existence of limitations on the scope of someone's reasonably held beliefs in such a way that they may not themselves reasonably be believed by that person. Of course, the one-day announcement is overtly bewildering to the pupils in a way that the many-day case is not; but that, on this account, would be a reflection only of the relative complexity of the argument needed to establish that the pupils cannot have reason to believe both A_n and C^*. In both cases the pupils have no right, on the strength of the announcement alone, to make any assumption about what the headmaster's intentions are.

The Quinean idea, then, generalised, is that $D_0^s A_n$ and $D_0^s C^*$ are capable of conjoint truth only on interpretations of the D-operators under which the the falsity of C^* is perfectly compatible with the headmaster's ability to surprise the pupils in an intuitively relevant sense. But a different account remains desirable for values of $n > 1$. Before we indicate how it might go, we had better try to diagnose the impression of many commentators that the puzzle is one of impredicativity.

IV

A clear disanalogy between the one- and many-day announcements is this: whereas the truth of the one-day announcement is inconsistent with the pupils merely having been given reason at 0 to believe A_1, there are no grounds for thinking that A_n and C^* cannot both be true provided that the pupils only have reason to believe A_n on Sunday and not also C^*. That is, the trouble in the many-day case would seem to have to do with the fact that the totality of information relative to which the examination is to come as a surprise includes C^* itself. This might suggest that Shaw,[3] Montague and Kaplan,[4] *et al.*, were right in their general contention that some sort of impredicativity is essentially involved in the mechanism of the paradox. But what is the truth of the matter?

In the presence of d(ii) a statement of the form, $\sim D_t^x p$, can always plausibly be argued to conceal a quantifier since it is equivalent to:

$$(\Gamma)[[\Gamma \text{ entails } p] \rightarrow \sim D_t^x \Gamma].$$

Thus C^*, it will be remembered, was preferred to the more cumbersome

$$C: \qquad (m)(\Gamma)[[\Gamma \text{ entails } E_m] \rightarrow (E_m \vee \rightarrow D_{m-1}^s \Gamma)].$$

Suppose that we had used C instead; what range would the natural deduction have required for the variable, Γ? The intuitive point, that after the announcement the examination has to be a surprise relative to a pool of information *including* the proposition that it will be a surprise—which is doubtless what has motivated the tradition of explicitly impredicative constructions of the announcement—might lead us to expect that Γ will be required to include C itself in its range. But this would be a confusion. Define a proposition as *elementary* if it is a truth-function solely of propositions of the form, E_m. Then C is non-elementary; but it will suffice for the purposes of the natural deduction if the range of Γ encompasses only elementary propositions. All that is needed of the surprise premise is that it carry us from E_{n-k} to $\sim D_{n-(k+1)}^s E_{n-k}$ at each $(k+1)$th stage of the reasoning as it works back through the week. So the only value of Γ which we have to consider at each stage is E_{n-k} itself—which is exactly what we did. Correspondingly, while C itself 'conceals' a quantifier equivalent to:

$$(m)(\Gamma)[[\Gamma \text{ entails } E_m] \rightarrow (\sim E_m \vee (\Delta)([\Delta \text{ entails } \Gamma] \rightarrow \sim D_{m-1}^s \Delta))],$$

all the natural deduction would require of the ranges of Δ *and* Γ would be that they include each E_{n-k}, *etc.*

The explanation of this is not difficult to see. Suppose that the above characterization of an elementary proposition is embedded within a theory of orders of the familiar sort, *i.e.* with orders defined according to levels of quantification. Then in the presence of d(ii) the notion of p's being logically independent of *all* propositions reasonably believed in advance, *i.e.* of its truth, if it is true, coming as an absolute surprize, is not stronger than that of its logical independence of all antecedently reasonably believed propositions of an order among which it has equivalents: in particular, of the order of p itself. Thus if the true E_m is unpredictable by means of any set of reasonably believed elementary propositions, the presence of d(ii) guarantees its unpredictability by means of any set of reasonably believed propositions whatever.

[3] 'The Paradox of the Unexpected Examination', *Mind* 67 (1958).
[4] 'A Paradox Regained', *Notre Dame Journal of Formal Logic* 3 (1960).

It is certainly part of the intuitive meaning of the headmaster's announcement that the examination will be a surprise *absolutely*, that the pupils will have no reason *whatever* to expect it on the day on which it actually occurs. And, certainly, if he can communicate this intention to them, it will enter into the corpus of information which, on his aim, will prove too slight to furnish the correct prediction. So it looks as though the surprise premise must involve appeal to a totality—*all* sets of propositions—of a sort redolent of paradox, and that the making of the announcement will result in a *de facto* violation of the 'Vicious Circle Principle'. But the plain fact is that we can handle the totalitarian aspect of the relevant notion of surprise without recourse to unrestricted quantification, still less the sort of overt self-reference inbuilt by some commentators, so long as an analogue of *d*(ii) is in play.

We suggest, therefore, that this puzzle is not one of impredicativity: at least if, as had been traditionally supposed, to define in predicative terms a notion impredicative in some formulations is to absolve it from suspicion.

V

Are the *d*-rules analytic of the notion of reasonable belief? We have argued the case for *d*(vi); and while it is fundamental to communal discourse that something of the sort should hold for *d*(i), the natural deduction made plain that what is needed for a potentially paradoxical outcome is the supposition that the pupils can have reason to believe the headmaster's announcement on Sunday, not the stronger assumption that he can bestow such reason to believe.

An affirmative answer seems unquestionable for *d*(ii) and *d*(iii). Good reason to believe is essentially distributive across logical consequence—that is why we are able to *argue* with each other—and to have verified a proposition is par excellence to have good reason to believe it.

But what about *d*(iv) and *d*(v)? There are persuasive, if not absolutely conclusive, reasons for regarding *d*(v) as valid for this interpretation. To attribute to x good reason to believe p at t is less than saying that he actually (reasonably) believes it. Rather, a claim is being made about x's particular state of information at that time: roughly, that he is in a position, if he is sufficiently straight-thinking, recollective, *etc.*, to provide a defence of a belief in p which is cogent for anyone who knows no more than he does. But if x, so idealized, is in a position to construct such a defence, then he is in a position to verify that such a defence can be constructed.

The argument turns on the decidability of reasonable belief. That is, if a state of information justifies a particular belief, it is always possible, at least in principle, to recognize as much. This cannot be a feature of any truth-entailing epistemic concept unless its range of application is restricted to decidable propositions; so it is not a feature of knowledge, as we ordinarily use the notion. To say that x has good reason to believe p at t is to say something decidable by x at t by *reflection* upon the character of his state of information. So it is to say something which, if true, that very state of information will provide good reason for him to believe.

On the other hand, d(iv) is, manifestly, *not* analytic of reasonable belief. Good reason to believe p may lapse as more information becomes available; or stronger reason to believe the contrary may emerge. *Prima facie,* it seems aimless to point this out. Surely d(iv) ought at least to be harmless? If one has good reason to believe p at t, then it ought to be unparadoxical to suppose that one will continue to do so for an arbitrary finite period into the future.

But it is not merely *that* assumption which d(iv), in its role as rule of derivation, enables the pupils to make. If we had dropped d(iv) in favour of, say, the assumption that everything which the pupils have good reason to believe at the start of the week will retain that status till its end, only the last day would be eliminable on the basis of the resulting pool of assumptions and the surviving d-rules. For the new assumption will enter into the set on which A_{n-1} depends; and it will then be required that the pupils should have reason to believe *it* on Thursday evening, *n—2*, if Friday is in turn to be eliminated. Further, in order to eliminate Thursday, they will need to assume that on *Wednesday* evening they will have reason to believe that on Thursday evening they will have reason to believe the new assumption. *Etc.*

We leave it to the reader to satisfy himself of the details. The essential point is that use of d(iv) as a *rule* in conjunction with the others permits the derivation from $D^s_0 p$ of any multiply D-prefixed statement, $D^s_r \ldots D^s_t p$, $0 < r < t$. Premises of this type, becoming progressively 'longer' as the reasoning works back through the week, will be seen to play an essential rôle if the natural deduction is developed without the use of d(iv), making any necessary assumptions *ad hoc*. But there is no reason to think that they *are* consequences of $D^s_0 A_n$ and $D^s_0 C^*$, once the invalidity of d(iv) is noted. The pupils unwittingly assume a burden of assumptions about what they will have reason to believe, what they will have reason to believe they will later have reason to believe, . . . *etc.*

No such assumptions are tacitly involved in the one-day case. In order to establish $\vdash \sim D^s_0\{A_1, C^*\}$, as the reader may verify, we need utilise only d(ii), d(v) and d(vi). So it appears that the invalid d(iv) is indeed the crux of the puzzle, and that the many-day announcement can be received by the pupils as credible information as condition E requires and—unlike the one-day announcement—as intuition dictates. Let us try to expand this a bit.

Suppose that the pupils have neither better reason to believe A_n than to believe C^*, nor conversely; their sole ground for believing either is the announcement. In that case they are not entitled to the simplest of the tacit assumptions on which their reasoning depends. For by means of the d-rules which we have argued to be analytic of reasonable belief—d(ii), d(iii), d(v) and d(vi)—it may straightforwardly be established that:

$$\vdash \sim A_{n-1} \rightarrow [D^s_{n-1}A_n \rightarrow \sim D^s_{n-1}C^*].$$

If Friday comes and no examination has taken place, they will not be entitled reasonably to believe *both* A_n and C^*. Hence the reasonable course, on the present supposition, will be to suspend belief in each. The announcement will have effected a *de facto* collapse into the (*ceteris paribus*) informationless one-day version; and the headmaster will be able to implement it on the last available day by trading on this fact.

'But what' it will be protested, 'if the pupils have *better* reason to believe A_n than to believe C^*?' (Suppose, *e.g.* that they previously checked that the pack contains the Ace of Spades.) In that case, consider their situation on Thursday evening, $n—2$, if the 'examination' has still not occurred. A Saturday examination would violate C^*; so if C^* is true, so is E_{n-1}. Can it in these circumstances still be reasonable for them to believe C^*? Supposing that their reason for believing A_n would in fact survive the non-ocurrence of the examination on the morrow, $n—1$, we have to consider two possibilities respectively: that they have, and that they have not, reason to believe on Thursday evening that this will be so.

Suppose, first, that they have reason to believe that tomorrow night they will still have reason to believe A_n, come the examination tomorrow or not. In that case, they can no longer reasonably believe C^*; that is, they cannot by the acceptable d-rules, as the reader may verify:

$$\vdash \sim A_{n-2} \rightarrow [[D^s_{n-2}A_n \ \& \ D^s_{n-2}D^s_{n-1}A_n] \rightarrow \sim D^s_{n-2}C^*].$$

Their reasons for believing C^* are thus discredited by the non-occurrence of the examination on or before Thursday. But now, *lacking* any reason to believe C^*, they have no reason to rule out a Saturday examination. Hence the way is open for the headmaster to fulfil the announcement by setting the examination on Friday.

Suppose, on the other hand, that while they will in fact continue to have reason to believe A_n on Friday evening, they lack on Thursday evening reason to believe that they will; *i.e.* $\sim D^s_{n-2}D^s_{n-1}A_n$. Then they have on Thursday evening no reason to believe that a Saturday examination would violate C^*, though in fact it would, nor therefore any reason to believe that if C^* is true, the examination will occur on Friday. Thus the examination can occur on Friday consistently both with its being a surprise and with the pupils' having good reason to believe both constituents of the announcement on Thursday evening, $n - 2$; contradiction is avoided because we lack $D^s_{n-2}D^s_{n-1}A_n$.

Lastly, suppose the pupils' reasons to believe C^* are stronger than those for believing A_n, *e.g.* they have overheard the headmaster confide to another member of staff that rather than set a non-surprise examination, he would abandon the whole idea. Then non-occurrence on or before Friday will rob them of reason to believe that there is still going to be an examination, and it will therefore be open to the headmaster to surprize them with one on Saturday.

What the pupils ought to realize from the start is that *one* way for the headmaster to carry out the announcement will be to let a situation develop in which their reasons for believing A_n and C^* are mutually discredited; or in which, if those reasons are of differing strengths, the weaker are discredited. Just as, by the announcement, he *prima facie* places A_n and C^* among the set of propositions which they have reason to believe, so by subsequent inaction he may take them away again. So while on Sunday they may regard themselves as entitled to believe A_n and C^*, they have no reason to assume that they will be so entitled at any stage in the week. Of course they will be so entitled in any period of the week remaining *after* the surprise examination; but it may be that the headmaster plans a *hiatus* between the period when their belief is justified by the announcement and the period when it is justified by memory.

The pupils should realize that there need be no such hiatus; it is merely that they have no reason to suppose that there will not be. Consider again the case where superior grounds are possessed for believing A_n, and where indeed the pupils have reason to suppose that they will continue to have such grounds throughout the week. In this case, as we noted, their reasons for believing C^* are discredited by the non-occurrence of the examination by Thursday night. But consider their situation on Wednesday night, $n—3$, if the examination has not so far occurred. If they had reason to believe that the headmaster intended *not* to let a situation develop in which their reasons to believe C^* would be discredited, they would have reason to rule out the non-occurrence of the examination by Thursday night, contrary to C^*. But if they lack any reason to think that they will continue to have reason to believe C^* throughout the period, they will have no reason to rule out a Friday examination; and it will then be consistent with their reasonably believing A_n and C^* at every time before Thursday that Thursday, $n—2$, should be examination day.

It appears, then, to be open to the pupils to accept the announcement as an informative expression of intention which they have every reason to believe. Of course, it will cease to be reasonable for them to take this view if the headmaster leaves things late enough—and they ought to recognize that possibility. But there seems, in general, no tension between one's having good reason to believe p and being forced to acknowledge that in certain envisageable circumstances this belief will no longer be possible, even though p may still be true.

If our analysis, then, does justice to the essential logic of the pupils' reasoning, it appears that all six of our original conditions can be met. When the D-operators are interpreted as expressive of reasonable belief, A_n and C^* conjointly capture the headmaster's intention to set an examination whose date the pupils will have no reason to predict: exactly the sense of 'surprise' which is intuitively relevant (condition C). But there is then no reason to think $\{D_0^s A_n, D_0^s C^*, A_n, C^*\}$ to be an inconsistent set (conditions A,B,E); the impression that it is so depends on play with suppressed premises, derivable from what is given only by recourse to a rule, d(iv), which is invalid for this interpretation of the D-operators. Condition F is met by the reflection that, even with d(iv), the natural deduction cannot proceed unless the person(s), k, to whom the announcement is directed, *i.e.* for whom it is assumed that $D_0^k A_n$, C^*, be the same as whoever is denoted by the person index of the D-operator *within* C^* itself. Finally (condition D) although there is no temptation, once one reflects on it, to think d(iv) valid for this interpretation or, indeed, to suppose that all the additional premises which the pupils will otherwise need, $D_{n-1}^s\{A_n, C^*\}$, $D_{n-2}^s D_{n-1}^s\{A_n, C^*\}$, *etc.*, *ought* to be consistent with the initial basis, the fact remains that the rôle of these premises in the informal reasoning is, apparently, not all that easy to detect.

VI

There are two principal reservations which it is natural to feel about this way of dealing with the puzzle. First, only one interpretation of the D-operators has been considered: in particular, we have said nothing about the situation if they are interpreted, as is more usual, in terms of knowledge. Second, no argument has been given for thinking that the puzzle depends upon implicit recourse to exactly d(iv); that is, may there not be some other rule, valid for reasonable belief, which will serve to generate trouble for our initial four assumptions?

Given the objective—to meet the six conditions, A-F—it is of no consequence that we have considered only one interpretation. The essential thing is that the interpretation be such that in order to meet condition E, *i.e.* to vindicate the intuitive impression that the pupils can learn from the announcement exactly what is announced, it suffices to establish the consistency of $D_0^s A_n$ with $D_0^s C^*$ under that interpretation. 'X is inclined to question p at t', for example, does not meet this requirement; but reasonable belief does, and that justifies the special attention which it has been given. Naturally under different interpretations which do meet this requirement we may wish to call into question d-rules other that d(iv); or perhaps each of d(ii)-d(vi) will be unquestionable, so that there will be no alternative to the conclusion that the pupils cannot . . . . . both A_n and C^* at time *0*. If that is the situation with some concepts of knowledge, then indeed the pupils cannot, in those senses, *know* that they are going to be set a surprise examination. The intuition that they ought to be able to learn the headmaster's intentions from his announcement will, nevertheless, be salvageable by the reflection, if we are right, that they can reasonably believe it, and so 'know' it, in a standard colloquial use of that term.

But of course the consistency of $D_0^s A_n$ with $D_0^s C^*$ under the reasonable belief interpretation has *not* been demonstrated. What has been shown is that the pupils' reasoning, as formalized in the natural deduction, provides no reason for thinking those statements inconsistent. Might there not be some *other* way of representing their reasoning, involving no use of d(iv), by means of which the initial assumptions would be validly shown to be inconsistent even when the D-operators were interpreted in terms of reasonable belief? If so, as we saw in section III, we shall after all be constrained to conclude that the pupils cannot have good reason to believe A_n and C^* on Sunday. Therefore, in advance of determining that no such alternative construction exists, we have no right to claim to have met condition E.

It is not clear how to allay a general doubt along these lines. But are any plausible alternative d-rules in sight? Binkley[5] considers:

$$d(\text{vii}): \quad \frac{\Gamma \vdash D_r^x \Delta}{\Gamma \vdash D_r^x D_t^x \Delta} \quad r < t.$$

To be sure, using only this rule and the d-rules argued to be analytic of reasonable belief, no way is apparent for deriving from its initial basis the additional premises which successive stages of our natural deduction will require. But what we can achieve is a proof at each kth stage not of A_{n-k} but of $D_0^s A_{n-k}$, depending only on $D_0^s A_n$ and $D_0^s C^*$ as assumptions. We leave it to the reader to check the details. So we shall have, finally, a proof of $D_0^s A_1$ on those assumptions; then d(ii), d(v) and d(vi) will swiftly yield the assumptionless conclusion that $D_0^s A_n$ and $D_0^s C^*$ are incapable of conjoint truth.

Binkley argues that d(vii) is valid for an 'ideal knower', ideal in respects which he details. But his argument for the principle takes account only of the special case, for which it certainly is valid, where x's experience at t will constitute a decision of the truth-value of p. The case relevant to present concerns is rather where r and t both occur too soon for us to be sure whether p is true or false, and where reasonable belief, not knowledge, is required. We are asking, then, whether if one has good reason to believe p at r, one *thereby* has good reason to believe that one will always be similarly placed at any subsequent t which is still too early for a verification or falsification of p. If we waive irrelevant complications to do with mortality, *etc.*, there seems to be nothing to be said for such a principle. Certain sorts of reasons to believe may be inherently lasting, but the natural view is that the principle is counter-exemplified precisely by this kind of case: for, to repeat, the pupils ought to recognize that one way for the headmaster to do what he has announced that he will do is to leave matters so late that their reasons to believe the announcement lapse; they have no reason to suppose that this will not be his strategy.

The last point suggests that if he attempts to enlarge the original

[5] 'The Surprise Examination in Modal Logic', *Journal of Philosophy* 65 (1968).

announcement so as to give the pupils reason to believe that this will *not* be his strategy, the resulting total statement will be, like the one-day case, incapable of being reasonably believed by them at time *0*. Actually, matters are a little more complicated. But rather than presume on the reader's curiosity, we leave him to satisfy it for himself.

We can conclude, at any rate, that the puzzle *may* have been disposed of, that no reason remains for thinking that our original six conditions cannot all be met. Whether there is a principle, perhaps after all *d*(vii), which is valid for reasonable belief and will reintroduce the problem is a question which awaits better grasp of the logical properties of reasonable belief. Since this notion is the sister of justified assertion which, rather than truth, may be the appropriate central notion in a general theory of meaning, there is a more powerful motive for securing this grasp than the desire to achieve final clarity about the Paradox of the Unexpected Examination.

All Souls' College, Oxford and *Monash University*

Received December 1976