The Surprise Examination in Modal Logic
Author(s): Robert Binkley
Source: *The Journal of Philosophy*, Vol. 65, No. 5, (Mar. 7, 1968), pp. 127-136
Published by: Journal of Philosophy, Inc.
Stable URL: http://www.jstor.org/stable/2024556
Accessed: 15/07/2008 12:32

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/page/info/about/policies/terms.jsp. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/action/showPublisher?publisherCode=jphil.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is a not-for-profit organization founded in 1995 to build trusted digital archives for scholarship. We work with the scholarly community to preserve their work and the materials they rely upon, and to build a common research platform that promotes the discovery and use of these resources. For more information about JSTOR, please contact support@jstor.org.

http://www.jstor.org

THE SURPRISE EXAMINATION IN MODAL LOGIC *

MODAL logic of an ordinary sort, when construed as a logic of belief or judgment, gives a schematic characterization of an ideally rational mind, or, as I shall call him here, an ideal knower.[1] It ought therefore to be of use in philosophy when we are dealing with problems which involve that ideal. One such problem is posed by the well-known paradox of the surprise examination, and in the present paper I attempt to diagnose that paradox with the aid of modal logic. I do this mainly in an attempt to vindicate the utility in philosophy of the logical machinery, but partly also because the paradox raises some important questions about the ideal knower's attitude toward his own future beliefs. And, of course, the paradox, though piddling, is not without a certain fascination of its own.

The paradox can be set out in the following story: A teacher announces to his class that there will be an examination in the afternoon of exactly one of the following n days, where n is some positive integer, and that the examination will take the students by surprise. The students protest that this announcement cannot be fulfilled, for if the examination is left until the last day the students will be able to anticipate it on the morning of that day; if that day is ruled out, the same reasoning will apply to the next earlier day, and so on until all the days are eliminated. But in spite of this, the announcement *is* fulfilled when the teacher gives the examination on one of the days (and why should he not?), thus catching the students by surprise.

Though the paradox and my treatment of it can be constructed for any n, it will suffice here to consider only the case when $n = 1$ (the one-day case), and the case when $n = 2$ (the two-day case).

That the paradox does involve the notion of an ideal knower has not always been remarked, but is easily seen. Clearly the puzzlement to which the story gives rise does concern minds (the students') and what they can know or judge, and equally clearly it does not concern either fortuitous obstacles they might encounter in acquiring knowledge (defective intelligence, lapse of memory, etc.) nor fortuitous aids (clairvoyance, teacher inadvertently revealing the day, etc.). What the teacher promises is that even an ideal knower placed in

* This paper is to some extent dependent on research supported by National Science Foundation grant GS-907.

[1] I have attempted to develop this idea, both for judgment and for practical decision, in "A Theory of Practical Reason," *Philosophical Review*, LXXIV, 4 (October 1965): 423–448.

the students' position and working with the data available to them would be surprised.

In an effort to abstract from all fortuitous elements of the situation, many writers offer a formulation in terms of deducibility. The teacher's announcement is construed in some such way as this: "There will be just one examination occurring on the afternoon of one of the next n days, and on the morning of no day can it be deduced from that fact, information about the examinationlessness of preceding days, and *what I am now saying,* that an examination will occur on that day." [2] As the italicized clause suggests, this leads to a diagnosis of the paradox in terms of self-reference. But it seems to me that such an approach misses the nub of the difficulty, which is that the flawless reasoning of the students is somehow rudely brought to nothing by the actual occurrence of the promised, but apparently impossible, examination. This intrigues us not just as an example of professorial one-up-manship, but because it has the "flavour of logic refuted by the world".[3] To paraphrase Bennett (*op. cit.*), *this* is the feature that makes the announcement teasing to everyone, and the puzzle it generates cannot be handled in a formulation that makes no use of epistemological or pragmatic concepts, as the deducibility formulation seeks to do.

We introduce such concepts when we invoke the ideal knower, to the more exact characterization of whom I now turn. First of all, it should be noted that we are concerned with an ideal seeker after knowledge, not necessarily someone who already possesses knowledge, and that consequently the ideal knower must be defined in terms of what he judges or believes, not in terms of what he knows.[4] He is meant to be an ideal of rationality, and circumstances may conspire to prevent a rational man from acquiring knowledge, and perhaps may even lead him into false belief.

Let us assume an ordered series of occasions on which judgment may take place (the mornings of the possible examination days), to be labeled 1, 2, ..., n. We introduce the operator 'J', which makes a sentence out of another sentence and an occasion label, so that '$J_i p$' says that on occasion i the person in question judged that p. Next, we construct a logical calculus, and define the ideal knower as someone of whom all the theorems of the calculus are true. The

[2] J. Bennett and J. Cargile survey the possibilities along these lines in a very elegant fashion in their reviews, *Journal of Symbolic Logic,* xxx, 1 (March 1965): 101–102, and 102–103.

[3] M. Scriven, "Paradoxical Announcements," *Mind,* lx, 239 (July 1951): 403–407.

[4] D. Kaplan and R. Montague, "A Paradox Regained," *Notre Dame Journal of Formal Logic,* i, 3 (July 1960): 79–90, explore that alternative approach.

calculus will be pretty much an ordinary modal logic, 'J', or rather 'J_i', functioning as the necessity operator. This is done by adding the following rule and the following axiom schemata to the propositional calculus:

R1. If α is a thesis, then $J_i\alpha$ is a thesis
A1. $J_i p \supset \sim J_i \sim p$
A2. $[J_i(p \supset q) \cdot J_i p] \supset J_i q$
A3. $J_i p \supset J_i J_i p$

All of these are familiar ingredients in logics of belief, and call for little comment. R1, A1, and A2 make the ideal knower a master logician; he avoids contradiction, is aware of all logical truth, and believes all the logical consequences of what he believes. R1 has the additional feature that it makes the ideal knower aware that he is an ideal knower, since, by R1, he must believe of himself all the axioms that define the ideal knower.

A3 adds self-knowledge to the other virtues. This again is not particularly controversial. Some writers, including myself but not Schick,[5] favor adding the converse of A3. In fact, if conditions are assumed guaranteeing that every case of not judging is a case of deliberately withholding judgment, it becomes plausible to add the still stronger axiom $\sim J_i p \supset J_i \sim J_i p$, which would yield a calculus resembling S5, but these additional strengths are not needed for dealing with our paradox. For this purpose we could even use a weakened form of A3, $J_i p \supset \sim J_i \sim J_i p$, which would be analogous to A5c discussed below.[6] But such weakening of A3 would be unreasonable. The very considerations about the ideal knower knowing his own mind which would be advanced to support the weakened form also support the stronger.

It must be noted, however, that A3 is a plausible axiom only if we presuppose, as we clearly may in the present context, that the knower always knows who he is and on what occasion he is judging. Otherwise, the knower, for example, might judge on Monday that the earth is round, but fail on Monday to judge that on Monday he judges that the earth is round because on Monday he thinks it is Tuesday. In a similar way, it will also simplify matters to assume, as I always shall, that the knower knows which occasions are earlier and later than which.

A very useful derivative rule of the calculus may be introduced at this point.

[5] "Consistency," *Philosophical Review*, LXXV, 4 (October 1966): 472.
[6] I owe this point to James Fulton.

DR1. If α follows from β (β and γ) in the system, then, $J_i\alpha$ follows from $J_i\beta$ ($J_i\beta$ and $J_i\gamma$) in the system.

This rule may be derived very quickly from R1. It gives formal expression to the point that the ideal knower judges all the logical consequences of what he judges. For present purposes, we could replace R1 by DR1, but this would give rise to a less familiar modal logic.

The machinery so far constructed is adequate to the one-day case, in which the teacher's announcement is, in effect "There will be an examination this afternoon that will take you by surprise." If we put this in symbols, letting e_i say that an examination occurs on the afternoon of day i, we get

$$e_1 \cdot \sim J_1 e_1$$

Statements of this form possess the peculiarity that, even if true, they cannot be believed by the relevant person on the relevant occasion, at least if that person is an ideal knower. We can show this in our case by demonstrating that if we combine the supposition that the students are ideal knowers with the supposition that on the first day they believe this announcement, then we are led to a contradiction. We suppose them to be ideal knowers by agreeing to use the present calculus about them; that they believe the announcement we take as a premise, and deduce the contradiction as follows:

1. $J_1(e_1 \cdot \sim J_1 e_1)$	Premise
2. $J_1 e_1$	1, DR1
3. $J_1 J_1 e_1$	2, A3
4. $\sim J_1 \sim J_1 e_1$	3, Al
5. $J_1 \sim J_1 e_1$	1, DR1, which contradicts 4

The students, therefore, cannot believe the whole of this announcement; they may or may not believe one or the other of its parts. If we suppose, as we may, that they do not believe the first part, that an examination is to be given, then the second part will be true. And if we suppose further that in fact an examination is given, then the first part will be true as well, and the teacher's announcement will be an incredible truth, a true proposition that the students cannot believe, even when it is told to them on the highest authority.[7] This is a sufficiently curious situation to deserve the name

[7] This point could also be put by saying that the students need to reckon with the possibility that what the teacher says may not be true, and that, consequently, in the one-day case as well as on the last day of the n-day case, they will not know ahead of time whether they are to have the examination or their teacher is to be proved a liar. Quine, in effect, points this out in "On a Supposed Paradox," *Mind*,

of paradox. It is, of course, very similar to what has become known as Moore's paradox, that is, a remark of the form "p, but I don't believe it." [8]

This diagnosis, I think, is sufficient for the one-day case; the interesting question is whether it can be extended to the two-day, and hence n-day, case. The answer will depend on what further assumptions we wish to make about the ideal knower, in particular on what interoccasional axioms we wish to lay down, for so far we have only considered the ideal knower on a single occasion.

One such axiom can be laid down at once. It concerns the special competence of the ideal knower when placed in the situation of the students. We must suppose that, so placed, the ideal knower would notice whether or not he was taking an examination, and would remember it in future. This, or rather the part of it that we will need, can be expressed as follows, where k is a later occasion than i:

A4. $\sim e_i \supset J_k \sim e_i$

A fuller account of the ideal knower would no doubt seek to derive this from more basic axioms together with a description of the students' situation, but this would not help with the present problem.

We now come to the final axiom, which concerns the relation in general between the ideal knower's judgments on different occasions. One such axiom, an ideal memory axiom to the effect that if something is judged on one occasion, then it is judged on later occasions that it was so judged on the earlier occasion, would perhaps be noncontroversial, but it is useless to us now. What we need of it is already included in A4. But there are three additional general interoccasional axioms that suggest themselves. Where k is a later occasion than i, we have:

A5a. $J_i p \supset J_k p$
A5b. $J_i p \supset J_i J_k p$
A5c. $J_i p \supset \sim J_i \sim J_k p$

The alternatives are given in order of decreasing strength; A5a implies A5b which implies A5c, but in no case do we have the reverse implication. And each one is deducible from the axioms we already have when k is the same as i.

Each requires a certain abstraction from real life. To be plausible

---

LXII, 1 (January 1953): 65–67, which has been reprinted as "On a Supposed Antimony" in his *The Ways of Paradox* (New York: Random House, 1966): 21–23. This places my own discussion in what might be called a Quinean tradition of skepticism with regard to authorities.

[8] See J. Hintikka, *Knowledge and Belief*, (Ithaca, N.Y.: Cornell, 1962), pp. 64–76.

at all, A5a requires that we abstract from the possibility that the ideal knower might lose his life, his memory, or his reason between occasions i and k, and the others require that the knower himself, at any rate, accept these abstractions. But these do not seem to me to be unreasonable idealizations, particularly in the context of our paradox.

Against A5a, it might be urged that it recommends that kind of stubborn mule-headedness castigated by Emerson in *Self-Reliance* in the following terms:

> A foolish consistency is the hobgoblin of little minds, adored by little statesmen and philosophers and divines. With consistency a great soul has simply nothing to do. He may as well concern himself with his shadow on the wall. Speak what you think now in hard words and tomorrow speak what tomorrow thinks in hard words again, though it contradict everything you said today.

It must be admitted that Emerson has something here. If it is pointed out to the great soul that his hard words of today contradict his hard words of yesterday, he will simply reply that he has changed his mind; today's evidence, which includes yesterday's evidence plus whatever new information has come in, indicates a different conclusion. What could be more reasonable?

But this consideration is not decisive against A5a. Against Emerson's great soul we must balance the Stoic wise man, of whom it is said:

> The Wise Man never opines, never regrets, never is mistaken, never changes his mind.[9]

Perhaps we may take this as the suggestion that the ideal knower will never *need* to change his mind because he will not speak any hard words at all until he has made absolutely certain that they will not need to be retracted. Taken in this spirit, A5a can be seen to represent not little-mindedness, but what we might call an idealized Cartesian epistemology, one which represents the accumulation of knowledge as the slow but sure building up of a structure, brick by solid brick, upon some secure foundation.

Considered as an idealization, this picture of the accumulation of knowledge is not entirely absurd, and may well have some application in special cases. But, taken as a general rule, it is too demanding. How could the ideal knower ever make his first judgment if he needed beforehand a guarantee of immunity from the need for fu-

[9] By Cicero following Zeno. Jason Saunders, *Greek and Roman Philosophy after Aristotle*, (New York: Free Press, 1966), p. 61.

ture revision? In fact, I think, we have all abandoned the Cartesian epistemological ideal and have become reconciled to a view of the knowledge enterprise as one in which false steps may occur. What we demand is that there be procedures that ensure that the false steps are eventually put right.

An epistemology that is thus willing to live dangerously must reject A5a, but it may still accept A5b. By doing this, we permit the ideal knower to make false steps, steps requiring subsequent retraction, and so he will realize in a general way that false steps for him are a possibility. But even so, it may be that, whenever he takes a step, he must think that *it* is not one of the false ones. This is A5b.

As it happens, A5a enters into the surprise-examination situation only as entailing A5b, and so it is unnecessary for us to choose between them. Rejecting A5b in favor of A5c, however, makes a considerable difference.

The appeal of A5c as against A5b can be appreciated in the following way. It seems clear that there is some kind of inconsistency in judging a thing today and at the same time judging that one will abandon the judgment tomorrow. (Remember that we are abstracting from such risks as the onset of insanity.) If one thinks that tomorrow's new evidence will overturn the judgment, why make it today? But, to avoid this inconsistency, it is enough to conform to A5c and suspend judgment today about one's judgment tomorrow. And what, one wants to ask, could be wrong with that? Why should we require the ideal knower to form a positive opinion about his future opinions? Once we recall the difference between positively judging that p and merely not judging that not-p, A5b will appear to be an unsatisfactory way station in the retreat from A5a. Let us simply speak today's hard words, and not worry about what we shall be saying tomorrow.

Now this line of thought is very tempting; but in spite of it I am going to defend A5b, though with a certain qualification. A5b, I claim, specifies the correct ideal *if* we are thinking of an ideal knower who uses his knowledge as a basis for planning for the future. Such a knower cannot suspend judgment about his future judgment. Suppose, to take a simple case, that I judge that it will rain tomorrow and, accordingly, form the plan to take an umbrella. What I do today is envisage a certain possible tomorrow; it will contain both rain and me with an umbrella. The question is whether this envisaged tomorrow *must* also contain me thinking that it is raining. Suppose that it need not, that is, that I may suspend judgment today as to whether I will still believe tomorrow in tomorrow's rain. This means that I may envisage a tomorrow in which I stand there with

my umbrella maintaining an agnosticism with respect to the weather, or even a positive belief in sunny skies. But this is absurd. And the reason for the absurdity lies in the fact that I would then be engaged in deliberate action, holding the umbrella, while no longer believing the reasons on which the decision to do the action is based. For a rational agent, when the beliefs upon which a plan rests go, the plan goes too; a new plan must be formed based on the new beliefs. From this I conclude that the kind of belief relevant to planning for the future is a kind that involves belief in future belief, in short, a kind conforming to A5b. And since in principle any belief may serve as a basis for planning, an ideal knower who is also a planner will have to conform to A5b. But for the benefit of those unconvinced by this argument, I shall also examine the paradox from the point of view of A5c.

If we accept A5b, then, in the two-day case, as in the one-day case, the teacher's announcement turns out to be an incredible though possibly true proposition. It now amounts to the following four assertions:

1. $\sim e_1 \supset e_2$
2. $e_2 \supset \sim e_1$
3. $e_1 \supset \sim J_1 e_1$
4. $e_2 \supset \sim J_2 e_2$

(1) and (2) say that exactly one examination will be given, while (3) and (4) say that it will be a surprise.

It is clear that these assertions can all be true together; the question is whether they can all be believed by the students, assuming that they are ideal knowers. That they cannot on the first day may be demonstrated as follows:

1′)	$J_1(\sim e_1 \supset e_2)$	
2′)	$J_1(e_2 \supset \sim e_1)$	
3′)	$J_1(e_1 \supset \sim J_1 e_1)$	
4′)	$J_1(e_2 \supset \sim J_2 e_2)$	
5)	$J_1(\sim e_1 \supset J_2 \sim e_1)$	A4, R1
6)	$J_1(e_2 \supset J_2 \sim e_1)$	2′, 5, DR1
7)	$J_1 J_2(\sim e_1 \supset e_2)$	1′, A5b
8)	$J_1(J_2 \sim e_1 \supset J_2 e_2)$	7, A2, DR1
9)	$J_1(e_2 \supset J_2 e_2)$	6, 8, DR1
10)	$J_1 \sim e_2$	4′, 9, DR1
11)	$J_1 e_1$	1′, 10, DR1
12)	$J_1 \sim J_1 e_1$	3′, 11, DR1
13)	$J_1 J_1 e_1$	11, A3
14)	$\sim J_1 \sim J_1 e_1$	13, A1

But (14) contradicts (12).

This demonstration, it seems to me, constitutes an entirely satisfactory diagnosis of the surprise-examination paradox, which is now seen to belong to the same family as Moore's paradox.

This demonstration can also be used to see what happens when A5b is rejected. If that axiom is replaced by A5c, then the present diagnosis of the paradox must be abandoned. But another diagnosis will become available, one which also relies on the paradoxical character of incredible but possibly true propositions, though in a less direct way.

Let us look at line (7), the only place where A5b is used in the demonstration. Without A5b, line (7) will not be forthcoming and the proof will collapse, unless, of course, some *ad hoc* means of supplying it is provided. Not only will (7) be missing; its negation can be deduced from our premises, since adding it to them leads to a contradiction. Now line (7) asserts the students' first-day belief in their second-day belief in one of the things the teacher said, namely, that there will be an examination on at least one of the days. What all this means, therefore, is that the students can believe the teacher on the first day only if they omit then to believe that they will still be believing him on the second day. In the absence of A5b, it is possible for them to do this without losing ideal-knower status. Moreover, if they do this their protest to the teacher evaporates. The examination may be given on the second day and take them by surprise because they have ceased by then to believe the teacher, and it may be given on the first, and surprise them because they had no reason to select that day for the examination rather than the second.

But why, believing the teacher on the first day, should they omit to believe that they will continue to believe him?

If the students believe what the teacher says at all, they must do so simply because the teacher said it, that is, because they trust him, for they have no other reason. But if they trust him enough to believe him on the first day, how can it be reasonable for them not to think that they will continue to trust him on the second? Clearly this combination of trust and doubt would be reasonable only if there were the possibility that some information might turn up between the first and second days that would cast doubt on the teacher's reliability. It is clear that some information might turn up by the *third* day to discredit him, for if no surprise examination has occurred by then, the teacher will have been proved a liar. And so it would be reasonable, perhaps, for the students not to believe on the first day that they will still be believing him on the third day. But this consideration does not apply to the second day.

What then can transpire between the first and second days to make trust in the teacher less reasonable? Given the idealizations we have assumed, the only relevant difference between the first and second days is that by the second day it will be known whether or not an examination occurred on the first. If an examination has been given on the first day, then the teacher's trustworthiness will be, if anything, enhanced, for his announcement will have come true, the students having had no reason to expect it then. But if an examination has not been given, then the teacher's trustworthiness will indeed be placed in question, and the students' caution confirmed; but in a curious way. It will not be because the teacher will have been proved a liar; a surprise examination may still be given. The reason is rather that events will have proved him to be a purveyor, by implication, of incredible propositions. Such people cannot be trusted. You cannot trust a man if he tells you things that you cannot believe.

This comes about because, if there is no examination on the first day, then the teacher's announcement becomes in effect the one-day paradoxical announcement considered above. For if we add $\sim e_1$ to what the teacher has said we can deduce $e_2 \cdot \sim J_2 e_2$, something that the students can't believe on the second day.

Now the students must envisage this paradoxical second-day test of their teacher's trustworthiness if their first-day doubt about their continued belief in his announcement is to be reasonably combined with their first-day trusting belief in it. But as we have seen, this doubt must be present if they are to believe it at all.

So either directly with A5b or indirectly without it, the surprise-examination paradox reduces to the phenomenon of incredible though possibly true propositions, and it should redound to the credit of modal logic that it helps us to see this.

ROBERT BINKLEY

University of Western Ontario

PRESUPPOSITION, IMPLICATION, AND SELF-REFERENCE *

THE two aims of this paper are, first, to explicate the semantic relation of *presupposition* among sentences, and, second, to employ the distinctions made in this explication in a discussion of certain paradoxes of self-reference. Section I will explore informally the distinction between presupposition and im-

* An earlier version of this paper was read at Duke University on May 12, 1967, and sections I and II were included in a paper presented at a symposium on free logic held at Michigan State University on June 9 and 10, 1967. Acknowledgments and bibliographical references have been collected in a note at the end.