THE SURPRISE TEST PARADOX

A COCKTAIL–PARTY version of the surprise-test paradox might run as follows: A teacher announces to his students that he is going to give just one test next week and that it will be a surprise, where, by 'surprise test', he means a test given on a day such that the students did not know by the night before that there would be a test on that day. The students reply, "You're wrong. Assuming we can depend on you to give a test, it won't be a surprise. It couldn't be a surprise Friday, because Thursday evening we'd know the test would have to be on the morrow. Since Friday is out, it couldn't be a surprise on Thursday either, because Wednesday night we'd know Thursday was the only day left...." The students continue in this way until every day is ruled out. Presumably, this is paradoxical because it seems plain that if the teacher does give a test, say, on Tuesday, the students really wouldn't have known the night before that this is what he was going to do, or at least, not without more to go on than the foregoing argument.

Some discussions of this paradox have come to the conclusion that the teacher's announcement really is false because, when made "precise," it comes to the self-referring contradiction:

A. (a) There will be exactly one test during school hours next week, and (b) the test will be held on a day such that it will not be possible to formally deduce, from premises consisting solely of this announcement (A) plus the premise that no test has occurred so far, which day the test will be.

Or, if it is preferred, we can have a self-referring paradox (B) by replacing 'formally deduce' in A, by 'formally deduce from true premises'.¹

A, and especially B, are of some logical interest, but they represent no new problem. If the teacher's announcement really did amount to one of them or something like them, then the surprise-test paradox would be nothing new. But the teacher's announcement is one that will be true if he gives a test on a day such that the students do not know by the night before which day the test will be, and false otherwise. And neither A nor B is like this.

Turning to the approach of showing that the student's argument is invalid, it has been observed that, whereas from "No test so far" on Thursday evening plus the teacher's announcement, "Test on

¹ For a more detailed discussion of A and B and some variations on them, the reader is referred to the reviews by Jonathan Bennett (101–102) and by the author (102–103), especially the one by Bennett, in The Journal of Symbolic Logic, xxx, 1 (March 1965): 101–103.
Friday” does follow, “The students know there will be a test on Friday” does not follow.²

It may be said that this is an extremely trivial objection to the students’ argument. But it is nonetheless worth having mentioned, since it helps to make clear what sort of argument it is. The conclusion that a Friday test would not do is not simply a matter of observing that the announcement, conjoined with “No test so far” on Thursday evening, entails “Test Friday.” It is rather a matter of showing that the students would know on Thursday if there were going to be a test Friday. And this involves assuming that Thursday night the students are going to be in reasonably good intellectual condition and that they have good reason to believe that the teacher will keep his word about giving a test some day of the week.

The students’ argument is thus not so much a matter of formal deduction as it is a matter of having good reasons for claiming to know. If the problem is to be at all interesting, we must allow the students to know there will be a test next week (suppose that the principal is known to have strictly ordered one) and to know that there will be no more than one test (again on principal’s orders) and to know that the teacher is a rational agent who will do his best to make his announcement true.

This latter assumption is important because the students’ justification for a claim to know the test will not be Friday must be that the teacher, who determines when the test takes place, will see that if he gives the test Friday it will fail to be a surprise in the appropriate sense. If a student had simply predicted a surprise test next week, no argument could get started against his prediction, because the power behind the test would have nothing at stake to lead it to avoid Friday.

A further assumption needed for interest is that the students should remain rational and efficient thinkers throughout the week. If the students knew they were going to remain drunk all next week, the announcement would be trivially true because the students wouldn’t be able to deduce anything.

With these assumptions, the students are right in concluding that there couldn’t be a surprise test on Friday. They could know perfectly well Thursday night that a test is coming for the next day, if there hadn’t been one so far.

At the next step, the students argue that there couldn’t be a surprise test Thursday, on the grounds that Wednesday evening they

would know the test was going to be on Thursday. What grounds can they advance for this claim?

It couldn't be that Thursday is the only day left for a test, for Friday would do equally well for a test. But Friday has been ruled out for a surprise test. And on the assumption that the agent determining the test day will do everything in his power to make the test a surprise in the present sense, it seems reasonable to infer that he will avoid Friday. And this leaves Thursday as the only day he can turn to to avoid Friday. Hence, on Wednesday evening the students would conclude the test will be Thursday.

This step in the argument is an interesting one deserving close attention, but let us pass it by without comment for now in order to note how the argument goes on. It has the form of an induction. Since Thursday and Friday are out for a surprise test, the students would know Tuesday evening the teacher will try Wednesday, and so on.

This induction seems ridiculous, and has, in fact, been one of the features of the argument commonly cited as paradoxical. For it would apply in the same form even if the teacher had announced a surprise test for the coming year, rather than the coming week.

Note that the induction has a form such that, before the week begins, the students will be picking Monday as the test day on the grounds that neither Tuesday nor Wednesday nor Thursday nor Friday would do for a surprise test, and the teacher will see this and thus be led to try Monday. But this is ludicrous. If the teacher is able to see the students' argument, he should simply respond by holding the test not Monday, but Tuesday. Then if the students try to claim to have known Monday evening the test would be Tuesday, the teacher can simply point out that similar grounds hadn't been very reliable with respect to Monday.

To this the students may reply that even the best of grounds may sometimes break down, and still suffice when they don't break down. But let us set this aside without discussion. There is a simpler way to show that the induction is fallacious.

Suppose that the teacher has announced a surprise test for the next year, and that this year consists of 300 school days. Then the teacher could respond to the induction with a method for choosing the test that would make the induction obviously inapplicable.

He could pick someone at random off the street, and ask him if he thinks zero persons is too many for a picnic where everyone gets to know everyone else. Then he asks whether one is too many, then two, etc. When he reaches a number \( n \) such that the subject answers yes,
the teacher makes the \((300 - n)\)th day the test day, or if \(n\) exceeds 300, holds the test the first day.

Now it is obvious the last day won't be reached, since the chances of a subject who is loose on the street thinking that zero people is too many for a cozy picnic are nil. But the students have no way of knowing (except irrelevant ways such as looking up the interviewee or breaking into the schedule, etc.) exactly what \(n\) is. They can rule out the last day, and the next to the last, and the third from last, and who knows how many more? But running the induction all the way would be to commit the fallacy of the beard.

It might be replied that the teacher might not think of the foregoing way of selecting the test day. This is quite true. The teacher might even be so confused by the inductive argument that he gives the test on the first day, so as to be as far away from those other days as possible. And the students might know the teacher is in this muddle. But then their claim to know, although valid, is based on special features of the teacher which make their success unparadoxical.

We have thus disposed of one paradoxical feature: the students' argument can't be validly extended indefinitely. And of course we needn't look for an exact point where it goes wrong. But now let us return to the second step in the argument, the step to the effect that on Wednesday night, the students would know the test will be on Thursday.

There is a selection procedure that the teacher could follow at this point which would make the students' argument clearly inapplicable. He could go to the principal Wednesday night and ask him to decide whether the test will be Thursday or Friday. (Here it is assumed that the principal knows nothing of the duel between the teacher and the students and that he is thoroughly authoritarian, so that the teacher will have to abide by his decision.) The students' grounds for expecting the test Thursday is that the person determining the test day will be avoiding Friday. But, with the principal determining the test day, these grounds are no good.

However, this selection procedure has a drawback not possessed by the selection procedure previously outlined for the 300-day case. For there is a solid chance that the principal will choose Friday, and, if he does, the students will know Thursday night there is going to be a test on the morrow, thus falsifying the teacher's announcement. So this selection procedure does not guarantee the teacher's announcement against falsification, but only leaves it a toss-up.

Now suppose that the teacher considers using this selection procedure and decides not to, on the grounds that he can give the test
Thursday and argue that the students' basis for claiming to have known Wednesday evening it would be Thursday is unsound, because, for all they know, he might have resorted to the random method of choosing the test day.

The students might reply that they knew him too well to overlook that and claim that they knew he wouldn't be able to resist foregoing the random method. Here again we have a claim that can't be assessed apart from considering the degree of special knowledge the students have of their teacher.

However, there is a problem here if we assume that both the teacher and the students are the sort of agents who figure in theorems in the theory of decision. Such theorems involve determining the optimum strategy, if any, for each player in some given game of strategy. And it is commonly assumed that the players are ideally rational and know each other to be such. This means at least that the players are assumed to reason just as well and ingenuously as the person proving the theorem does. A strategy that is good against a formidable opponent may be bad against a trifling one, since it may involve measures that are inconvenient and unnecessary in the latter case. Part of the theory of decision is concerned with strategies suitable against formidable opponents, and the notion of an ideally rational opponent is supposed to capture this notion of formidableness.

In taking up the problem in these terms, let us consider some passages from a noted text on decision theory:

As we said in the last section, the players in the game are to be characterized by three assumptions, only one of which has so far been given, namely: each player has preferences over the outcomes which meet the axioms of utility theory. The other two assumptions concern what the players know and the basis on which they arrive at decisions.

Let us take up the question of knowledge first. It is assumed that:

viii. Each player is fully cognizant of the game in extensive form, i.e., he is fully aware of the rules of the game and the utility functions of each of the players.

. . . this is a serious idealization which only rarely is met in actual situations . . . [but] the theory can be used normatively to tell a person that this is the knowledge he should acquire, and once he has it, the theory establishes the decisions he should make in order to achieve certain specified ends. . . .

Finally, we must consider a "law of behavior" for the players. . . . We then postulate that

ix. Of two alternatives which give rise to outcomes, a player will choose the one which yields the more preferred outcome, or, more precisely,
in terms of the utility function he will attempt to maximize expected utility.\footnote{R. Duncan Luce and Howard Raiffa, \textit{Games and Decisions} (New York: Wiley, 1957), pp. 47-48.}

Before taking up the surprise test, it may be worth setting out an example to show how these strong assumptions may be employed. Consider the following case:

There are ten philosophy departments having equal prestige, and each is an ideally rational Bayesian agent whose primary value is prestige, and this is common knowledge to each department. The departments do not communicate and are subject to no higher authority. Van Jones, a Harvard graduate student of world renown, is about to take a Ph.D. and is looking for a job. He has confided that he will confine his choice to one of the ten departments and will choose randomly among any offers he receives from them.

Inviting Jones and getting him will yield 7 prestige points. Inviting him and not getting him will yield a loss of $-5$ prestige points (because the world in general isn't going to be told Van Jones chose at random between the offers). Not inviting him and pretending not to have been interested all along will yield 0 prestige points. If \textbf{I} represents inviting Jones and \textbf{J} represents getting Jones, the probability matrix is:

\begin{align*}
 & \textbf{J} & \sim \textbf{J} \\
 \textbf{I} & p & 1-p \\
\sim \textbf{I} & 0 & 1
\end{align*}

and the desirability matrix is:

\begin{align*}
 & \textbf{J} & \sim \textbf{J} \\
 \textbf{I} & 7 & -5 \\
\sim \textbf{I} & 0 & 0
\end{align*}

Combining these in the usual way, we see that the expected utility of not inviting Jones is zero and the expected utility of inviting Jones is $12p - 5$, where $p$ is the probability of his accepting if invited. So if $p$ is over $5/12$, a department will decide to invite him.

At first, one might think: "If it is reasonable to invite Jones on the data available, each department will do so, since they all have the same data and preferences. But this would make $p$ below $5/12$ (it would be $1/10$). And if it isn't reasonable none would do so, so $p$ would be above $5/12$ (namely 1). So it is reasonable to invite Jones if and only if it isn't."
However, this would be to overlook the fact that a decision other than simply choosing to invite or not to invite can be dictated. Being ideally rational, it would occur to each department to select the following strategy: Each department tosses a coin in a series so that a certain outcome $A$ has a probability $X$, such that if the departments followed the strategy of tossing for outcome $A$ and invited Jones only if they got $A$, then the probability of getting Jones if you are authorized (by having gotten $A$) to invite him is above $5/12$, and, further, $X$ is such that if you didn’t get outcome $A$ and decided to cheat by inviting Jones anyway, then your cheating would lower below $5/12$ the probability that inviters of Jones will get him.

This strategy, if adopted, would itself forestall cheating despite the absence of an overseer. And it seems plausible to expect ideally rational agents who don’t miss bets to hit upon it.

This example is of interest because it shows how simply having the data that your competitors are ideally rational and know you are and know you know they are can lead to a decision almost out of nothing. We now need to examine the effect of such strong assumptions when introduced into the surprise-test problem.

What is given is that the teacher has to give a test, and strongly prefers its being a surprise in the present sense to its not being a surprise. And both he and the students are ideally rational—“will not miss a bet”—and know each other to be so. Let us suppose that the desirability of giving a test that is a surprise is 1 and the desirability of giving a test that is not a surprise is $-1$ and the probability of giving a test that is a surprise on Friday is 0. Then the expected utility of giving the test on Friday is $-1$. In order for a Thursday test to be the rational decision for an ideally rational Bayesian agent, the expected utility of a Thursday test must exceed $-1$.

This may be seen by setting up matrices in the usual way, with ‘$T$’ for “test on Thursday,” ‘$F$’ for “test on Friday,” and ‘$S$’ for “the test is a surprise.” Then the teacher’s desirability matrix is:

\[
\begin{array}{c|cc}
 & S & \sim S \\
\hline
T & 1 & -1 \\
F & 1 & -1
\end{array}
\]

and his probability matrix is:

\[
\begin{array}{c|cc}
 & S & \sim S \\
\hline
T & \rho & 1-\rho \\
F & 0 & 1
\end{array}
\]
Combining these gives the result that the expected utility of a Friday test is $-1$, and the expected utility of a Thursday test is $2p - 1$, where $p$ is the probability of a Thursday test's being a surprise. So a Thursday test will be *dictated* by Bayesian considerations to our ideally rational teacher if and only if the probability $p$ of such a test's being a surprise is nonzero.

Here, by "the probability $p$" we mean something determined by the teacher and not some "objective" figure. In fact, there may be no figure involved at all. The point is just that the teacher knows there is no chance (zero probability, if you please) that a Friday test will be a surprise. And from this plus our idealizing assumptions it follows that if the teacher thinks there is any chance at all (nonzero probability, if you don't mind, otherwise proceed in terms of "chance") of a Thursday test's being a surprise, then his ideal rational Bayesianhood will dictate that he choose Thursday. On the other hand, if the teacher thinks there is no chance at all for a surprise test Thursday, then, since he also knows there is no chance for a surprise test Friday, his ideal rational Bayesianhood will leave the decision arbitrary.

So the crucial question for our teacher is whether there is any chance of a Thursday test's being a surprise. His answer to this question will determine the nature of his decision. If his answer is yes, he will have to choose Thursday. If his answer is no, then his decision will be open to any arbitrary selection procedure.

Could his answer be "I don't know"? Perhaps. But, for the purpose of deciding, this would be the same as "Yes." If one is obliged to choose between $A$ and $B$ and knows only that $B$ is not preferable to $A$, then mustn't $A$ be the rational choice?

Whether or not that is a rule of decision which a rational agent must have, the following would appear to be an essential truth about ideal rationality: If two ideally rational agents are asking independently whether a given proposition is true and if both have exactly the same relevant data and exactly the same knowledge about what is relevant, then they will both reach the same conclusion. The conclusion may be "Yes" or "No" or "insufficient data to determine" or "the question is unclear," etc., but it must be the same for both. For suppose that the two agents arrive at different answers, $X$ and $Y$. Then $X$ cannot be a better answer than $Y$ on the information given, since that would contradict the assumption that both agents are ideally rational—that is, think as well as is possible in every case. But then the answer "$X$ is no better an answer than $Y$" is determinable on the information given and is clearly a better answer than $X"
or $Y$, which contradicts the assumption that both agents will give the best possible answer on the information available to them.

This principle concerning ideally rational agents leads to a paradox in the present case. For in considering the question whether there is a chance of a Thursday test's being a surprise, our ideally rational teacher and students have exactly the same relevant information and knowledge about what is relevant. So if the teacher thinks Thursday has a chance, the students will know he thinks this because they will have reached the same conclusion themselves and they will know the principle concerning rational agents. And if he thinks there is no chance for Thursday, the students will again know he thinks this, having gotten the same conclusion. And, if the teacher concludes, "I just can't tell whether there is a chance or not," the students will know he thinks this because they will have thought the same thing. And even if the rule "If obliged to choose between $A$ and $B$ knowing only that $B$ is not preferable to $A$, then choose $A$" does not apply to ideally rational agents (and I think it should), then there will be some course set forth for an ideally rational agent for the "I just can't tell" answer, and the ideally rational students will know it just as well as the ideally rational teacher. Specifically, the course would probably be "choose at random" (not "choose as you like," since preferences have presumably been taken into account in arriving at the toss-up), though I do not exclude other possibilities, such as "toss a coin at a ratio of $m/n$," for some value of $m/n$.

Suppose now that it is a school rule (known to both students and teachers) that the principal must be notified at exactly 10:00 P.M. on the evening preceding any day a test is given, and the teacher finds himself, at 9:00 P.M. Wednesday evening, trying to decide whether to notify the principal in order to hold a Thursday test.

We have already seen that if the teacher's preferences and ideal rationality dictate a choice of Thursday, the students will know this as well as he. This means that Thursday can't be the ideally rational choice. But the only way the students can defeat the teacher's announcement by knowing the test will be Thursday solely on the basis of knowing the teacher's ideally rational Bayesian nature and his preferences, is for Thursday to be the ideally rational choice. So if Thursday isn't the ideally rational choice, then if the test happens Thursday, the teacher will win. And the teacher can see this as well as we can. But surely if the teacher knows a Thursday test will win, then Thursday is the ideally rational choice for him. So we seem to have the result that Thursday is the ideally rational choice if and only if it is not. And this contradiction seems to follow naturally enough
and from plausible enough assumptions to deserve the title of "the surprise-test paradox"; furthermore, this seems to be a new sort of paradox, rather than a version of a paradox of the self-referring sort.

Here it is worth warning against the tendency to say that the announcement, by virtue of being self-refuting in some sense, thus becomes true. In this case it might be said that, since the students will run up against a paradox in the course of their ratiocinations, they will be without a basis for knowing in advance and Thursday will be ideal. The answer to this is that we ran into the paradox just as squarely as the students did (and so did the teacher). So if we can see Thursday is ideal, so can the students, and we are back into the paradox. We have to resolve it, rather than try to add it as an ingredient to an even wilder story. We have enough trouble already.

It will also simplify considerations to add to our case the specification that the teacher has to choose the test day directly by himself, and that this is common knowledge and known to be common knowledge. That is, the principal will not select the test day, and there is no neutral party who will select the day for the teacher. Thus, if he tosses a coin to decide, he will still have to decide to stick by the coin's advice. There is no means available for randomizing his choice. He might make the odds as steeply in favor of Thursday as you please, but the fact remains that the students know that if the outcome should happen to point to Friday, the teacher will be subjected to the question whether to shake off the signal; and so the paradox cannot be avoided by saying that it is just a surprising proof that the teacher should choose randomly—it is perfectly possible that he should not have the facilities (such as a principal willing to choose and enforce the choice) to do this.

It might be thought that the teacher could draw the test day from a box full of envelopes with a mixture of "Thursday"'s and "Friday"'s, and hand the envelope directly to the principal at 10:00 P.M. But we require that the principal will accept only a direct verbal statement from the teacher, and the teacher has no means (such as drugs, etc.) of making this statement unconsciously.

In turning now to the problem of resolving the surprise-test paradox in the terms just presented, we may begin by noting that the blame must lie with the idealizing assumptions about rational Bayesian agents. For the arguments for the paradoxical conclusion could not stand serious scrutiny in the absence of this idealization. And yet, once the teacher is portrayed as the sort who always makes the best choice on the given data (if there is a best) and this is assumed to be common knowledge, and the students are credited with perfect
rationality, the arguments gain sufficient plausibility to justify talk of paradox.

In view of this, it would be natural to dismiss the paradox by simply rejecting these assumptions. Of course this need not mean rejecting them in their entirety or foregoing all the uses to which they have been put, but it is not easy to isolate the "safe" part of the assumptions. And both the assumption that there can be agents who can reach the best possible conclusion in given circumstances and the assumption that they may know each other to be such are used heavily in obtaining numerous plausible results.

Before dropping the assumptions, let us consider the situation in retrospect. It is Saturday, and the teacher did not lose his wits to the extent of throwing away his job by disobeying the principal's order for a test. The question of whether the teacher or the students won is brought before a perfectly rational judge, who, incidentally, can depend on the two parties to be perfectly candid about what they thought.

First suppose the test was Friday. Then the students win. They simply point out to the judge that they knew the principal had ordered a test and they knew Thursday evening that giving a test Friday was the only way the teacher could obey this order.

Now suppose that the test was Thursday. One thing we just can't say is that the students win. Of course, they could have won if they knew something about the teacher's personality that would justify a claim to have known his motivation in advance. But in the present case, all the students know is that the teacher is perfectly rational and knows they are and prefers that the test be a surprise in the present sense and has to give a test Thursday or Friday. It may be a mistake to grant that they know this much. That is, our contradiction may lie in granting them this much. But they certainly can't appeal to more than this to justify their claim to know. And how could they claim to have known the test would be Thursday on the grounds that the teacher must have thought Thursday to be his best choice? Since all they know is that the teacher is ideally rational (if they know that), they can justify a claim that the teacher would consider Thursday his best choice only on the grounds that an ideally rational agent would consider Thursday his best choice on the given data. But such an agent couldn't reach such a conclusion on the given data without crediting the students with seeing it, too, and, if the students see it, it can't be the best choice any more, since it won't win.

The students (or rather some less rational spokesman) might reply,
"Yes, but there's all this paradox associated with Thursday, so that it is clearly a better day than Friday, just because the teacher can hope to get away in a cloud of dust." Again, how can the students blandly credit the teacher with being acute enough to see this without crediting him with being acute enough to credit them with seeing this?

Assuming that the fact that the students cannot win means that the teacher wins if he holds the test Thursday, then does this mean the ideally rational thing for the teacher to do is to choose Thursday? For example, suppose it is Wednesday evening and the teacher is pacing the floor in his room (which is just across the hall from the principal's) trying to decide what to do. The hour is drawing near to 10:00 P.M. and the teacher thinks, "Ah, Devil take it! I'll toss a coin, there's no chance of winning," and just because it lands for Thursday, he stalks over and opts for Thursday with the principal. Can we say that, in thinking in this way and risking a test Friday and winning only because the coin happened for Thursday, the teacher was thinking at a level less than the best?

It seems to me that we can say this, and can say that the ideally rational conclusion for the teacher is that the students can't win if he holds the test Thursday, so that the ideally rational decision for him is to hold the test Thursday. Here of course our paradox will be raised again. If it is ideally rational to conclude that the students can't win, mustn't the teacher see that the students will think of this? So mustn't he see that they can say, "We knew the test would be Thursday, because we saw that Thursday would win for the teacher and we knew that he would see this"?

Well, they obviously can't say this (so that it stands up) because the students can't justify having stopped with this. But while they can't justify stopping at this conclusion, it is nonetheless the right conclusion, and the teacher can stop at it. All he has to observe is that the students can't reach the same conclusion and justify stopping there.

This is certainly paradoxical, and "Surprise Test Paradox" is a fair title; but it isn't necessary to regard it as a contradiction in the assumptions about ideally rational agents. It is a contradiction only if we retain the "theorem" offered earlier, that: If two ideally rational agents are asking independently whether a given proposition is true, and both have exactly the same relevant data and exactly the same knowledge about what is relevant, then they will both reach the same conclusion. We may reject this on the grounds that the situation the agent is in matters in determining what he is justified in concluding.
The students should just admit they don’t know on Wednesday evening (though they would on Thursday evening if the test isn’t held Thursday) and the teacher must choose Thursday if he is smart.

Alternatively, someone wanting to retain the assumptions could just bar their application in any “games” in which the sole stake is knowledge, as opposed to getting a job, winning a race, etc. Behind this might be the feeling that the paradox is the result of a too abstract notion of knowledge which is too unrelated to practical courses of action and their consequences. This restriction on the application of the idealizing assumptions is indeed a third possible line of solution, in addition to rejecting the assumptions or the line considered in the preceding paragraph. However, the motivation for it can be taken up into the solution of the preceding paragraph.

That solution, that the students really don’t know on Thursday evening, essentially involved the assumption that the students are bringing their knowledge claim before an ideally rational judge. Such a judge will demand a high standard of argument, and won’t be satisfied with the student’s claim to know, since the claim will be based on holding that the teacher must have seen they wouldn’t know. It isn’t very good justification for a claim to point out that an ideally rational agent would have seen it to be false.

On the other hand, it would seem that the students could stand back from their situation Wednesday evening, and coolly survey all the possibilities, just as we have done, and reach the same conclusion, since they are very acute. And couldn’t they thus know pretty well that Thursday is the best day, and be pretty justifiably confident that the teacher will see this and be led to pick Thursday?

Well, we have seen that they can’t claim a Thursday choice is perfectly certain on grounds that will stand up before an ideally rational judge, but recognizing this is perfectly compatible with acknowledging that the students could be justifiably confident there will be a test Thursday. For example, they might justifiably lay a small side bet on a Thursday test, and qualify as making a wise bet. They may be perfectly justified in telling other classes they are going to have a Thursday test. If these school-mates ask “How do you know?”, the students do not owe them a full-dress reply. Confidence that is justifiable before one tribunal may not be justifiable before another, either because the standards are different or because the importance of the claim is different in each case.

For example, suppose that the University has asked me to notify them if I become certain that Jones is not going to pass my course. They want to know in advance if possible so they could cut him from
a list of national scholarship applicants which they must forward before graduation.

Here 'certain' is not made precise, but it does have rough boundaries. The administration is not interested in mere premonitions. They want to be fair, but if Jones is sure to fail, it will help them to know before they forward the scholarship list.

Now suppose that one day Jones gives a particularly shabby performance in class discussion. That night I say to my wife “Jones hasn't a chance of passing.” She asks “How do you know?” I say “Well, he's hard-working, but after talking to him at length I can see he just doesn't have the intelligence to master the material we have coming up.” This is enough justification. When Jones fails, I can remind my wife that I knew he wouldn't make it.

However, this needn't mean that I am in a position to make the same claim to the University as I did to my wife. There may still be a possibility that Jones will make it, which, though not a live possibility against the standards appropriate to after-dinner conversation about students, is a live possibility against the standard of evidence required for me to make a claim that will close a door to Jones, as opposed to one which merely influences my wife's opinion of Jones.

In this case it is a difference in what is at stake that makes the standards different in the two situations. In the surprise-test case it is not a matter of importance that makes the standards different, since it has been assumed that the students have nothing at stake with the teacher except a claim to have shown him wrong in a clever way. The situation is more like the sort that arises in connection with mathematical theorems. For example, Gödel might well have been credited with knowing that the axiom of choice is independent of the other axioms of set theory even before this was strictly proved. But while he might have claimed to know this before an informal panel of judges, the claim would have been unjustified before judges wanting detailed proofs, where coming before this panel is to put yourself forward as having such a proof. (Of course the degree of detail that might be required to count as proof varies considerably.)

This latter case is similar to that of the surprise test. When it is said that the students do not know, this must be taken with reference to an ideally rational judge who requires meticulous support for the claim. They certainly have ample justification for taking studying Wednesday evening to be a wise course of action.

JAMES CARGILE

University of Virginia