[SQUEAKING]

[RUSTLING]

[CLICKING]

IAN BALL: So the last topic of the course that we'll discuss today is going to be something called common knowledge. So throughout the course, we've often said things like, oh, we've assumed that the game is common knowledge. But we've never really been very precise about exactly what that means when we say the game itself is common knowledge.

So today, we're going to introduce formally this concept of common knowledge. And this is a topic that has quite a long history and has been studied across many fields. So philosophers thought about this issue. Logicians and mathematicians thought about it. And then economists have thought about it as well.

So the starting point of our model of common knowledge is we're going to say there's some set, this is called omega, of states of the world. And there's various interpretations about what these states of the world are. You could imagine every state specifying every possible detail about the world, every arrangement of atoms. Of course, that's not a very useful model.

So often, when we talk about states of the world, we talk about the states of the world, the dimensions or aspects of the world that are relevant to the problem at hand. And I think this is all a bit abstract. So as I introduce the model, I want to, on the other side, introduce some leading examples to make this concrete.

So as a leading example let's imagine we're learning about the roll of a die. And therefore, the state of the world is just going to be these possible values. Of course, this is a huge simplification, but I think this is going to be a good way to fix ideas.

So a die is rolled. It can take any value between 1 and 6. And it's also helpful to represent these states kind of geometrically here. So the exact arrangement doesn't matter. It just makes it easier to draw the pictures. So I'm going to think of the states as being like this-- 1, 2, 3, 4, 5, 6.

And then once we have the set of states the world, we can then talk about each player's-- it's called their information partition. So we're going to say an agent, say agent i, has information partition Ki. So the subscript i indicates this is something-- this represents the information for player i. And the K comes from the word knowledge, which we'll talk about later on. We'll talk about what knowledge is.

Now, partition is just a mathy word that says we're arranging these different states, these different dots, and we're splitting them up into groups. That's what a partition is. It just puts things into groups. So in this opening example, let's imagine the groups are like this.

So this is an example of one possible information partition for player i. We've partitioned the six states into four different groups. And each of these groups, they're sometimes called a cell. So we might say that this partition assigns the states into one of four cells.

The first cell contains states 1 and 2. This cell contains only state 3. This cell contains only state 5. And these cells contain states 4 and 6. And the word "partition" may remind you when we talked about extensive form games, we talked about a player's information partition. And this is the same concept here.

So what's the interpretation of this? The interpretation is that a player, player i, does not learn the actual state. They don't actually see the roll of the die. But what they do learn is which cell the state lies in.

So in this case, the player might learn the state of the world is in this cell. And if they learn that, then they can say, well, I know the die was either 1 or 2, but I can't distinguish between 1 or 2. And similarly, but in this state, when the player learns that the state is in this cell of the partition, they know for certain that the state was 3.

Now, it's convenient to then introduce or denote Ki of omega to be the cell of the partition that contains state omega. This is just a definition that's going to make it a bit easier to reason about these things. So this is going to be informally the cell of the partition containing omega.

So again, I think it's easier to just see an example. Let's look at our example over here. What would Ki of, say, 4 be? Well, Ki of 4 is going to be the cell of the partition that contains the state 4. In this case, that's going to be this cell.

And we need to name this cell. Well, this is the cell that contains states 4 and 6. So we're going to write here this equals 4, 6. So what we're saying here is when the state is 4, the cell that contains 4 is the cell 4 comma 6.

And then as just one more example to make sure we're clear, we could say Ki of 3-- well, the cell containing 3 is just this cell that only contains 3. So this is going to be just 3. Any questions about just this notation so far?

So now we can start talking about knowledge. So what does it mean to know something? Well, the player doesn't necessarily know exactly what the state is, but they may be able to say that the state satisfies certain properties, or the state lies in some smaller set. And that's exactly how we're going to define knowledge.

I guess maybe the most difficult thing when we first think about knowledge is, what do you form knowledge about? When I say I know something, what does that mean? And the key point is knowledge is going to be formed about what we call an event.

And we can just think of this intuitively. I might say, I know that the die is odd. And if I say that, I'm making a statement about the event that the state is either 1, 3, or 5, the possible odd realizations.

So first, we're going to define an event. And this notation may be familiar to you if you've taken a class on probability theory. An event E is just a subset of the state space omega. So as an example, e.g., the event E equals 1, 3, 5, formally, it's a subset of the state space. But colloquially we might say this is the event that the die is on, that the state or the die is on.

So now that we have the first piece, we can talk about does a player know E? Do they know that the die is odd? But we need one more piece of notation here because, well, whether I know the die is odd is going to depend on what the actual value of the die is.

If the die is 6, I'm not going to know that the die is odd. So our definition is going to describe knowledge in a particular state. So we're going to say in state omega, player i knows E.

So notice here that our definition depends on two or maybe three things. It depends on what the actual state is. That's certainly affects knowledge. It depends on which player we're talking about because some players know things that other players don't. And it depends on the fact or the statement or the proposition that we're actually knowing, this event E. So we say in state omega player knows E if-- well, we're going to just write it in math and then interpret it, Ki omega is a subset of E.

Let's interpret this. If the state is omega, well, the player is informed by their information partition not necessarily the state is actually omega, but they're informed of which cell of the partition the state lies in. So all the player learns, all the player, quote unquote, "knows," is that the state lies in this cell of their partition.

And if that cell is a subset of E, then the player can conclude that the event must be E. They don't know exactly which state has been realized, but they know whichever state has been realized, it must be in here. And therefore, it must be in here. And this is exactly how we're going to define knowledge. So I think it's kind of easier to go through an example again.

So let's here think of an event, the event that the die is, at most, 4. OK. So that's going to be this event here. And we can ask, well, does this player, the player i whose information partition is represented here, do they know event E? Or another way we might say it is, do they know that event E has occurred?

Well, it's certainly going to depend on what the die actually was. So let's ask if the die is 5, does the player know E? Do they know that the event E has occurred? No, because this cell is not a subset of 5. And, in fact, there's only going to be three states-- this one, this one, and this one-- where the player knows the event E

So let's write that down over here as an example. So I guess here I've changed my event E. Let's call it the event E prime. So let's consider the event E prime, which is 1, 2, 3, 4. So this is the event that the die is, at most, 4. And let's just go through each of our cases.

So if the state is either 1 or 2-- let's start if the state is 1. If the state is 1, well, the player knows that the state lies in this cell 1, 2. And because the set 1, 2 is a subset of the set E prime, we conclude by our definition that in state 1 the player knows the event is E.

So we have Ki of 1 is a subset of E prime, which means player i knows E prime in this state 1. And we could reason analogously about state 2. And we could reason analogously about state 3. Now let's worry about state 4. I think that's maybe a tricky case.

So state in state 4, the cell of the partition for player i is the cell 4, 6. So in this case, Ki of 4 is not a subset of E prime. Oops. Why? Well, this cell, 4, 6, is not entirely contained within the yellow box. It breaks the yellow box. And parts of this cell lie outside the yellow box. So we conclude that in state 4, player i does not know E prime.

Now, notice the difference here. When the state is 1, the event E prime has occurred, and the player knows that it has occurred. In the state 4, event E prime has occurred because 4 is in E prime. Yet, the player does not know this.

And then we can have an even easier case. Let's look at Ki of 5. And again, Ki of 5 is not a subset of E prime. So here, player i does not know E prime.

But notice there's kind of an interesting difference here. In state 4, E prime has occurred, but the player i doesn't know it. In state 5, E prime has not occurred, and the player doesn't know it. And this is consistent with the way that we drew information partitions and thought about extensive form games previously.

So I guess the first thing we might ask, well, we saw three possibilities here. E prime occurs, and we know it. E prime occurs, and we don't know it. And E prime doesn't occur, and we don't know it.

You might ask, is it ever possible that I know as an event when the event doesn't happen? And it turns out that's going to be impossible, and we're going to make that precise next. So I guess the next definition we need-- well, we've said what it means for player i to know an event in a given state. But now we can talk about well, in which states does a player know a given event E? And that itself is another event.

So now the event that player i knows E is given by-- well, we're going to call it Ki of E. So this is going to be the set of all states in which player i knows that event E has occurred. So let's write this down mathematically. It's the state omega.

First, maybe we'll write it in words, that at omega player i knows E. So this is itself an event because it's a collection of states. And it's the collection of all states, such in that state player i knows that event E has occurred.

And now we can just plug in our definition from here to make this more precise. And we can say, this is the set of omega omega, such that-- so all I've done is I've replaced this informal statement that player i knows E to a more precise statement of what that knowledge means, according to this definition over here.

And now we can try to formalize this property that a player can only know something if it's true, that if in some state I know an event E holds, then it must be the case that the event E actually holds. And let's state that mathematically and then prove it.

So here's our maybe first very basic result. Ki of E is a subset of E. So let's understand what this means. This says in any state of the world in which player i knows that E has occurred, it must be the case that that state is actually in E. And therefore, the event E has in fact occurred.

And the proof of this is actually-- I mean, I think conceptually it may be hard, but mathematically pretty straightforward. We want to argue that this set is a subset of this set. So we're going to take an arbitrary thing in this set and argue that it must be in this set. So let's choose some state omega that's in Ki of E. And now let's just apply our definition.

So what does it mean that state omega is in Ki of E? If we apply our definition, what can we conclude about omega? We just want to go to our definition. Yeah, Amy?

AUDIENCE: [INAUDIBLE]

IAN BALL: Right. So this is just by definition, right? If we choose omega in here, well, by definition, that means Ki of omega is a subset of E. We haven't done much here. We've just used our definition. And then I would argue we're basically done.

What's the last thing we need to see when you want to argue that, in fact, omega is in E? So what do we need to do? Yeah?

AUDIENCE: Omega [INAUDIBLE] of Ki.

IAN BALL: Yeah. So we're basically done. Ki omega is the cell of the partition that contains omega. So, in particular, omega is in here. So what that tells us is that omega is in Ki of omega, which is a subset of E. And therefore, omega is in E.

Maybe that's just a mathier way. We can look at it over here. If I'm in some state, and the cell containing that state is entirely contained in this yellow box, then the state must be in the yellow box. That's all we're saying.

And this is kind of a fundamental property of knowledge. I should maybe point out an alternative approach to modeling knowledge is not to work with a primitive partition, but to start with this function that tells us which events do we know, and then imposing axioms or properties on this function. And if you impose enough properties on this function, you can show that it must be derived from some partition in this way. But we're not going to go through this entire exercise.

So now we've just talked about things that a single player knows. Now we're going to talk about multiple players at the same time. So let's come back over here, do a really simple example.

So now let's consider we have two players. We have players one and two, and each player may know different things about the state. And therefore, each player is going to have their own information partition. So player one is going to have their information partition K1, and player two is going to have their information partition K2.

And let's represent the state as we did before. So this is state 1, 2, 3, 4, 5, 6. And I think it's easier now if we use colors. So let's put player one's information partition in blue. And let's say that player one's information partition looks like this. And now let's say that player two's information partition looks like this.

So intuitively, what does player one learn? Player one learns either that the die is 1 or 2, or that the die is 3 or 4, or that the die is 5 or 6. What does player two learn? They learn either that the die is odd or that the die is even. So now we can talk about the different information that the players have.

And now let's draw an event here. So let's let E be the event 1, 2, 3, 4, 5. So now maybe first, let's ask, what is the-- let's describe the event K1 of E. So this is going to be in which states does player one know that event E has occurred? Any ideas here? What is this going to be Yeah, Amy?

AUDIENCE: 1, 2, 3, 4.

IAN BALL: Yeah. And actually, we can just write them all together. Yeah. So 1, 2, 3, 4. In any of these states-- so that's an important point. The object here, it's not a collection of cells. It's a collection of states. So it is the states in this cell and the states in this cell. So altogether, we get 1, 2, 3 4.

And then we can ask, what about player two? Well, similarly player two, it's only going to be 1, 3, and 5 because these are the only states that are in a cell that's contained in the yellow box. So we get 1, 3, 5.

But now we can ask another question, which is, what is K1 of E intersect K2 of E? So let's try to interpret this set. This is the set of states where player one knows E, and player two knows E. And if we just read it off here, we can see these are the states 1 and 3.

So we can say in state 3, player one knows E has occurred, and player two knows E has occurred. But now we can start asking maybe a more confusing question. Does player one know that both players know that event E has occurred? So let's try to think through what that would mean.

So we want to ask-- we want to look at the set K1 of K1 of E intersect K2 of E. So this means I'm player one. In which states do I know that I know event E, and my opponent, player two, knows events E? So now we can start thinking about interactive knowledge. What do I know about what other players know?

And it's kind of a mouthful to say it, but mathematically it's not too hard. We've just said what this is. This is 1, 3. So what is K1 of 1, 3? Yeah?

AUDIENCE: Would it be the empty set?

IAN BALL: It's just the empty set because there are no states in which player one's cell of his partition is a subset of 1, 3. So this is one possibility. It can be the case-- so we can say informally, what does this mean? It means there are states of the world where both players 1 and 2 know that the event is E.

But there is no state of the world where player one knows that both players know that the event is E. And in fact, this motivates one more definition. Let's define-- and this is what we might call mutual knowledge.

So it gets a bit messy to keep talking about, do I know that player one knows something and player two knows something and player three knows something. So let's just introduce notation for the event that everyone knows that some event has occurred.

So we'll say in state omega, event E is mutual knowledge if-- well, Ki omega is a subset of E for all i, for all players i that we're talking about. Maybe we'll say i equals 1 through n, because we often work with n players.

And then once we have this definition, we can just naturally define-- well, here I've said, in a given state an event is mutual knowledge if everyone knows it. But now we can ask about the event that everyone knows something, or the event that this event is mutual knowledge. I know it's getting to be a bit of a mouthful, but we can give another definition, and we'll just say K of E is equal to the intersection-- or maybe I'll write it this way. K1 of E intersect Kn of E.

So this is precisely the set of states where player one knows event E. Player two knows event E. And everyone else knows event E. So if we want to rewrite this equation star over here using our new notation, we could write that is the event K1 of K of E.

So in the notation, E is an event. K of E is the event that E is mutual knowledge. It's the event that everyone knows E is true. And K1 of K of E is the event that player one knows that everyone knows that the event E is true.

So what this highlights is that mutual knowledge sounds pretty strong, but it's actually not quite as strong as we might think. Because it may be that we all know something, but there's a player who doesn't know that we all know something. And that suggests that we might need to go a step deeper.

So we could then look at, what is the interpretation k squared of E is equal to K of K of E? So in words, what would you describe this as? This is the event that what? K of K of E. Yeah?

AUDIENCE: Everyone knows that everyone else knows.

IAN BALL: Knows that E is true. So we might call-- we call this mutual knowledge. It's kind of first-order mutual knowledge. This we might call second-order mutual knowledge. Everyone knows something. And on top of that, everyone knows that everyone knows it.

So this is maybe second-order mutual knowledge, but that's still too weak. Because it could be that everyone knows something, and everyone knows that everyone knows something. But maybe people don't know that everyone knows that everyone knows it. So have to go a step deeper.

And we finally get to maybe what I'll say common knowledge, CK of E. Now we can formally define this thing that we've been informally talking about throughout the course. CK of E is K of E intersect K squared of E Intersect K cubed of E, and it goes on forever.

So now we have a formal definition of common knowledge. Event E is common knowledge. Or in other words, the set of states in which event E is common knowledge are precisely the states in which everyone knows E. Everyone knows that everyone knows E. Everyone knows that everyone knows that everyone knows E. And this goes on no matter how long the length of this chain is. And this is a very, very demanding property that we've implicitly been kind of relying on throughout the course.

Now, I think you might think, well, why are we going through all this? This is kind of-- we're just playing math games. Does this really have any content? So I now, for the next part of the class, want to go over a famous riddle that illustrates how different common knowledge of something can be from just mutual knowledge, or from first- or second-order mutual knowledge So this is kind of a classic riddle that we'll go over.

So this is sometimes called-- I don't know. It has many names-- the sage and the children. I don't know. Or the teacher-- there's so many stories we could use. Let's use the teacher and the students. And here's the-- the story, of course, it's kind of a bit of a silly puzzle, but I think it illustrates the ideas really well.

So we have three students. Call them 1, 2, and 3. And there's various versions of it. We'll use the version of hats. So each student is wearing a hat. The student's hat is white or red.

Why do we use hats? Well, the point is, when you're wearing a hat, you can see other people's hats, but not your own. So you can observe the color of everyone else's hat, but you can't observe the color of your own hat. And so we'll say that each student observes the other's hats, but not her own.

So let's suppose that all the hats are actually red. We should put it in red. So all hats are red. This is an assumption. It's really a statement about what the true state omega is. So the true state is that all the hats are red. And then the teacher is going to ask all the students, do you know what color your hat is?

Now, for this riddle to work out, it must be the case that we're assuming everyone is truthful. No one's lying to the teacher. So the teacher asks, do you know your hat color? And we should be a little precise when it means know your hat color means either you know hat is red, or you know your hat is white. That's what it means to know your hat color.

What would the student's answer be if they're truthful? They look around. Do they know their hat color? No, right? They can see other people's hats, but they can't see their own. They know others hat colors. They don't know their own. So do your own hat color? The answer is no.

Then the teacher makes an announcement. The teacher says-- so the teacher announces at least one hat is red. Have people heard of this riddle before? I'm curious. Anyone? One person, two. OK, so not too many.

So the teacher says at least one hat is red. And now she's going to go through and ask these questions again to the students one by one. So she's going to go to student 1, S1, and say, do you know your hat color? And still, I mean, knowing that at least one hat is red doesn't give that student any information about their own hat. So that student is going to say no.

And then the teacher goes to student 2 and says, do you know your hat color? And the student says, no. And crucially, when the teacher is asking these students, the other students get to observe the answers. So these are public questioning, and we're assuming everyone's truthful. And then she goes to student three, and that student says yes. Now I know my hat color.

Now, I think people-- this, I think, was first developed in the 1950s. These mathematicians were really shocked by it. So why is this kind of shocking? I hope people are shocked.

It's shocking-- well, it doesn't feel like the teacher's announcement should have made any difference. The teacher announced something that everyone knew. The teacher said at least one hat his red.

Well, if everyone's had his red, any student can look around and, in fact, see that there's at least two red hats. So why would the announcement that at least one hat is red change the dynamics that come after? Any thoughts?

The teacher has said something that everyone knows. So why does this announcement make a difference using the language-- maybe using the language we've said before, how could saying something, announcing something that everyone already knows make a difference? Yeah?

AUDIENCE: If player one-- or student one and student two, both sides, one hat is red. And both, one that they don't share [INAUDIBLE].

IAN BALL: I didn't follow.

AUDIENCE: [INAUDIBLE] based on one student [INAUDIBLE].

IAN BALL: So student three is learning from based on what student one and student said. But let's be clear. If the teacher had not made this announcement, and the teacher kept saying, do you know your hat color? Everyone would say no, and we would just go around forever, actually.

So the answers to the questions are fundamentally changed by this announcement that at least one hat is red. So why does this announcement make a difference? What is changed by this announcement using maybe a hint about what we've talked about over here. Yeah?

AUDIENCE: So like everyone already knew that at least one hat is red, but not everyone knew that everyone knew it.

IAN BALL: Yeah. In fact, everyone knew that everyone knew it. But it's not the case that everyone knew that everyone knew that everyone knew it. So why did this announcement make a difference? It took something that was mutual knowledge-- in fact, if we're careful, it's second-order mutual knowledge, but it was not common knowledge, and it made that common knowledge.

So announcements about things that everyone know can change things. If that thing that everyone knows is only mutual knowledge and not common knowledge. So the crucial thing about this announcement is it makes an event go from mutual knowledge, everyone knows it, to common knowledge. And I think this really highlights how fundamentally different these two concepts are. So let's try to work through and see how we actually get this. And then we'll conclude.

So first, we have to model the state space. So each of the three students has a hat, and the hat can either be white or red. I find it easier to use binary here. So let's say we're going to say white is going to correspond to 0, and red is going to correspond to 1. So what is a state omega?

Well, a state omega is going to be something like 010. This would mean player one's hat is white, player two's hat is red, and player three's hat is white. So I'm going to represent the state by 3 binary digits, each one indicating whether that corresponding student's hat is red. So I'll just note here, this is student two's hat color.

Now, this means that we have eight states. And it turns out to be quite useful to represent the states as the corners of a cube. So let me think about different axes here. So we have the first axis, the second axis, and the third axis. This will kind of test my drawing abilities here. I think-- how do I do this? So we need- have I drawn that right? It keeps-- kind of jumps at you.

Let me see how I did it here. OK Yeah, let me draw this a bit differently. I don't know if it makes more sense to me or it will make more sense to other people, too. I think I should-- yeah.

I'm struggling here. I should have done better in drawing class here. Let's see. It makes sense in my paper, but then I can't draw it here. OK, I think it should be-- yeah. Yeah. There we go. That looks better.

So we have-- so this is state 111. This is state 000. And as we move on this axis, we increase component 1 by one. As we move along this axis, we increase component 2 by one. And as we move along this axis, we increase component 3 by one. So my picture-- I think-- there we go.

So now let's think about the information partitions. So what is player one's information partition? Well, player one's partition has four cells. The player one observes the colors of the other two students' hats, but they don't observe their own hat. So we can think of each cell of the partition as being a horizontal line.

And it may get a bit messy when I keep drawing everyone else's. Player two's is going to be like this. And then player three's partition is going to be like this. It's going to be [INAUDIBLE]. So let's understand what that means.

In this cell of the partition, player one sees that the other two player's hats are white, but player one doesn't know her own hat color. She knows her hat can either be white or red.

Let's go up to this cell. Well, this point here is 001. So here, player one sees that player two's hat is white and player three's hat is red, but she doesn't know her own hat color. And that's why these two are going to be like this. So this is-- any questions about this information structure? Great.

So now let me try to formalize my claim. My claim here was that the event, that at least one hat is read, is mutual knowledge but not common knowledge. So let's try to check that. So let's write this down. This event E-- maybe I'll write E greater than or equal to 1. This is the event that at least one student's hat is red. I'll write this as a triple.

So the event that at least one student's hat is red is the event that omega 1 plus omega 2 plus omega 3 is greater than or equal to 1. In fact, if we draw it on the graph, this is the event that's everything except this point. So it's the event containing all seven points of the cube, which I've quite messily drawn, but hopefully it looks-- you can see it now, except this point here.

So now we're going to have to do a bit of calculation. So let's come over here. So now let's ask, what is K1 of E greater than 1? So this is, in which states does player one know that at least one player's hat is red?

I want to be clear here. Over here, I was talking about the true state being red, red, red. But for these statements about knowledge, they don't depend on what the true state is. These are just statements about in which state something is known. They don't depend on the realized state. So what is this going to be? It's omega and omega such that what? Yeah?

AUDIENCE: Omega 2 plus omega 3 is greater than omega 4.

IAN BALL: Exactly. So I only know at least someone's hat is red if I see a red hat. And I'm player one. I see hats two and three. I look at them. If either of those is red, then I know at least one person's hat is red. So this is the event that W2 plus W3 is greater than or equal to 1.

And symmetrically, for player 2 it's the same thing, but it's W1 plus W3 is greater than or equal to 1. Because if I'm player two, I can only observe the hats of player one and player three. And then finally, K3.

So now that we've done these calculations, we can say, well, what is K of E, which, remember, is the intersection K1 of E intersect K2 of E. Well, W2 plus W3 has to be at least 1. W1 plus W3 has to be at least 1. And W1 plus W2 has to be at least 1. It turns out this is exactly going to be the event that at least two hats are red. Let's see how we can see this.

I think it's kind of easier maybe to reason in words. If only one person's hat is red, then the person with the red hat sees two white hats. So they're not certain that at least one person's hat is red. But if at least two hats are red, then everyone must see at least one red hat, and therefore, know that at least one hat is red. So formally, if we maintain this notation here, this is E greater than or equal to 2.

So let's see if we can continue this pattern. So we've said that K of E greater than or equal to 1 is E greater than or equal to 2. The event that everyone knows at least one hat is red is exactly the event that at least two hats are red. You need one more red hat for people to know that at least one hat is red.

And now let's ask, what is K squared of E greater than or equal to 1? This is the event that everyone knows that everyone knows that at least one hat is red. Well, from what we said over here, this is K of E greater than or equal to 2. This means everyone knows that at least two hats are red. And any guesses? Can we just reason through in which states does everyone know that at least two of the three hats are red? Yeah?

AUDIENCE: [INAUDIBLE]

IAN BALL: All three hats are red. So this is going to be E3. Because if anyone's hat was white, then a person seeing that white hat would only be seeing one red hat, and they wouldn't be certain that there are at least two red hats.

So now, final step. What is K3 of E greater than or equal to 1? This is K of E3. So in which states does everyone know that there are three red hats? Yeah?

AUDIENCE: [INAUDIBLE] red hats [INAUDIBLE].

IAN BALL: Let's think. So do we know-- let's say everyone has three red hats. Do we all know that everyone has three red hats? No. So actually, this is the empty set. Because even if everyone has three red hats, we don't all know that. Because I only know the hats of the other two people.

So here, we can formalize our claim in state omega equals 111. Well, the event at least one person's hat is red is mutual knowledge. It's second-order mutual knowledge, but it's not third-order mutual knowledge. So in this state, the event E1 is maybe I'll say second, but not third-order mutual knowledge.

And I think the reason this puzzle is so confusing to people is that we have pretty good intuition about mutual knowledge. We have pretty weak intuition about second-order mutual knowledge. And once we get to third- or fourth-order mutual knowledge, it's really hard for us to reason through it. It feels a lot like common knowledge.

So in this game, because this event is second-order mutual knowledge, it feels a lot like common knowledge. But in fact, it's not actually third-order mutual knowledge, and therefore, certainly is not common knowledge. And that's why this announcement can make such a big difference.

Let me pause to see if there's any questions on this, and then we'll see if we can actually resolve the riddle. Questions? Yeah?

AUDIENCE: Like, I guess if we were to continue, does that mean like the other students would also start to [INAUDIBLE]?

IAN BALL: You're saying if we ask more questions here?

AUDIENCE: Yeah. [INAUDIBLE].

IAN BALL: If we ask a fourth question.

AUDIENCE: If we just keep going around.

IAN BALL: Yeah, it turns out it would actually cycle. Then it gets even more confusing, and it's going to depend on what the state is. But you can actually have situations where it would cycle where you might think, oh, we're just learning more and more information.

But yeah you can actually have this kind of weird cycling pattern where the subsequent answers now change, and we keep going on forever. Yeah. Any other questions? Yes?

AUDIENCE: Do you mean-- [INAUDIBLE] answers [INAUDIBLE] that should be-- yes. [INAUDIBLE] students actually don't know their hat color, or a student [INAUDIBLE].

IAN BALL: So you're right. So if I know my hat color, I can never subsequently forget my hat color. That's true. So what am I thinking in this example? Let me-- I should double-check this.

Yeah. So more information can ever cause you to not know something that you previously knew. But there is a similar riddle where these answers can cycle. So they must be asking about do you know something about-- it must be something about something other people knowing something about you.

Let me follow up with that. Yeah, let me double-check that. You're right, that more information-- adding information can never cause you to forget something you once knew.

So now let's see if we can actually resolve the puzzle and see what's happening here. So I think it's best to go to the picture here and go through the questions. So here we are. The teacher, I guess, first makes an announcement. I'll go here-- so these stages.

The teacher makes an announcement. Basically, they say, the event is E greater than or equal to 1. The teacher says at least one person's hat is red. But crucially, this announcement makes this event common knowledge. And what it means is now everyone knows that everyone knows that everyone knows that we're not in this state. So let me put an x over this state.

So now we're going to go to player one, and we're going to ask player one, do you know your hat color? That is, do you know that your hat is white? Or do you know that you had his red? Or are you not sure? And let's go through each of player one's information sets.

So if the state of the world is in this information set, does player one know their hat color? No. Because player one thinks either they're in this state, or they're in this state, and they don't know their hat color. If they're in this information set, player one also doesn't know their hat color, and same with this information set.

But if player one is in this information set, well, now, because we've announced that the state is not 000. Basically, what's happened is this information set has split into two. The partition has changed.

And in fact, if we're really formal about it, that's true for this one and also for this one. It gets messy to keep writing this. But the announcement that the state is not 000 splits up any of the information sets, any of the cells that includes this state 000. So it splits up this one, this one, and this one.

So now when we ask player one, do you know your hat color, and player one says no, what do all the other players now learn about the state? What is the one state that they can rule out? Yeah?

AUDIENCE: The bottom one.

IAN BALL: This one, right? Because if the state were here, player one would know for certain that his hat was red. So now the players-- we break this up. And I want to be really clear here, what these questions are doing is they're changing the information partitions that the players have. Again, this one is going to split up, and this one is going to split up. Because now all the players learn that we're not here.

Now let's go to player two. So next, player two is asked, do your hat color? Well, let's look. If the state is in this information set or this cell for player two, then player two doesn't know their hat color. If it's in this cell, then player two also doesn't know their hat color. But if it's in either of these four points, the player does know their hat color.

Because now these four points are in distinct singleton information sets for player two. Because previously, these two were in the same information set in the same cell, but that was split by the first announcement. And then these two were in the same cell, but that was split by player one's answer. So when player two says no, I don't know my hat color, that tells us that the state is not here or here.

So now everyone knows that the state must be up in this higher horizontal plane. Player one knows we're either in this information set or this information set, and they actually know which one they're in. Player two sees this. But player one now, player three actually now perfectly knows the state. Why?

Because if the state is here, then player three knows it. If the state is here, player three knows it. If the state is here, player three knows it. And if the state is here, player three knows it. So now player three, when the state is actually 111, is able to conclude that the state must be here. And therefore, the answer is yes.

I think this can be a little confusing. So yeah?

AUDIENCE: Please repeat what you said about after player's two answer, how that allows player one to know a state?

IAN BALL: Yes. So before player two answers-- I think maybe another way of looking at it. Let's focus on player three. So maybe another way of looking at it without breaking information sets is player three's information set is this one.

Because the state is 111, we're saying suppose the realized state is that everyone's hat is red. So that means that player three knows either my hat is red and the other two player's hats are red, or my hat is white, and the other two player's hats are red. Put this up here. These are the only two possibilities that player three entertains because they can observe the other two player's hats.

Now player three now says, well, wait a second. Suppose the state were this. If the state were 110, and that means that my hat is white, but the other two player's hats are red, then let me put myself in the shoes of player two. So if this were the state, then player two's information set would be this.

But player two has learned from player one's answer that this can't be the state. So if this were the state, then player two would know for certain that this were the state. And therefore, player two would have said yes to this question.

So it must be the case because player two says no, that we can't be in this state. And therefore, we must be in this state. And knowing that we're in this state, I know that my hat is red.

Great. So I think I want to end a bit early since it's the last class. I'll stop there. Thanks.

[APPLAUSE]