

[SQUEAKING]

[RUSTLING]

[CLICKING]

**IAN BALL:**

All right. So today, we're starting a new phase of the course, which is going to be about games with incomplete information. And I'll warn you up front that the terminology here is a bit confusing. We've been talking about imperfect and perfect information, and now we're talking about complete and incomplete information. Complete and perfect sound similar colloquially, but in game theory, they have very different meanings. So I want to be precise about what's going on.

So far, we've always looked at complete information. And what's confusing is that under complete information, under this complete information umbrella, there are two special cases that we've talked about or two distinct cases, which is perfect information and imperfect information. So I'll give some formal definitions, but then I'll also give some examples to keep in mind, because I think these definitions can seem a bit abstract. So in games of perfect information, what that meant was that every information set was a singleton.

And I think the canonical game to keep in mind for games of perfect information would be games like chess. So when we play chess, whenever it's our turn to move, we can perfectly see how everyone else has moved so far in chess. We see the position of the board, and we remember each move that everyone else played. That's the sense in which every information set is a singleton.

Then we also looked at games of imperfect information. What that meant was that some information sets were not singletons. So I'll say some non-singleton info sets. Now, economically, what does it mean when we had non-singleton information sets? It meant that there were times when someone was called upon to move, and there was some aspect of the game or of the previous moves they didn't know about-- either moved by nature or moved by a different player. And they knew which information set they were at, but they didn't know which node within that information set they were at.

And this was the case-- there's a few leading examples here. Whenever we had a simultaneous move game-- whenever the players moved simultaneously, remember, the way we represented that in the extensive form was to say, well, technically one player has moved first. But then the other player moves without observing the move by the other player, and therefore that second player's information set is going to include different nodes that specify how the other player moved.

So simultaneous moves-- the simple example here, I would say, is poker. Poker would be another case, because when I'm called upon to move in poker, I don't know the cards that the other players have been dealt. Even if we move sequentially in poker, my information set is still non-singleton because when I move, I don't know what cards the other player has. So these are two distinct ways we can have imperfect information.

But why do I say both of these are complete information? So the defining feature, the defining property of these two classes of games is what makes them complete information. And I'll warn you, some people give slightly different definitions of this, and I think my definition is slightly different than what's given in the notes. But what I like to think of it is, at the beginning of the game, information is symmetric.

So beginning of the game is crucial here. Once we start playing poker, we're going to be dealt hands. And one player won't know what cards the other player has been dealt. But when we start the game, when we sit down at the poker table, whatever one player knows, everyone else knows. No one's been dealt any cards yet.

We all understand the rules of the game. We all understand that people want to win as much money as possible, and information is symmetric in the sense that anything known by one player is known by all players. And maybe if I want to be more precise, I'll say it's common knowledge.

This doesn't mean we know everything. When we sit down at a poker table, we don't know what hand we're going to be dealt in the future. But none of us know that. There's no information that one player knows that other players don't know at the beginning of the game. And it's crucial here that we're at the beginning of the game.

The distinction with incomplete information-- when I say information is symmetric, let me add here, relevant information. Of course, when you sit down at a poker table, you know your birthday and someone else doesn't. But that's not really relevant to the game. So in a technical sense, we never have complete information because there's always these unmodeled things that are irrelevant. But what I mean is relevant information with incomplete information.

The difference is that at the beginning of the game, there is relevant information that some player knows but another player doesn't know. So there is information, or more precisely, relevant information that is known by some player, but not others.

And let's give a few examples of this. So examples of incomplete information games. And what you'll see is that a lot of our examples of complete information games come from parlor games. They come from things like chess or poker. But in the real world, when we're analyzing economic situations, we're much more often in the incomplete information case. So let's give a few examples of that.

Was there a question? Yeah.

**AUDIENCE:** No, nevermind.

**IAN BALL:** You're welcome to ask.

**AUDIENCE:** Oh, I thought you were asking for example.

**IAN BALL:** Oh, sure. Give an example, yeah. Sorry, sometimes I'm rhetorical. It's not clear. You can always answer any question you want. And if I didn't intend it, that's totally fine. Yeah, go ahead.

**AUDIENCE:** Would one of them be when there's price haggling and they don't know each other's values?

**IAN BALL:** Exactly. That's a great example. And actually, that fits in. The first example I was going to give is a special case of that. So let me give this is the kind of classic example. And it exactly fits into your setting. And this, I would say is buying a used car. And of course, price haggling over the used car.

Let's say we're haggling over the price of the used car. What don't we know? Well, you brought up one example. We don't know each player's valuation for the car. And we can be a little more specific, right? Let's say a salesperson is selling you a used car. There's something they know that you don't, probably. What might they know that you don't know about the car? Yes, in the second row?

**AUDIENCE:** The condition?

**IAN BALL:** The condition of the car. So in this case, the seller, maybe they're an experienced mechanic or auto dealer. They've seen a lot of cars. They've examined the car closely. You maybe don't know a lot about cars. So often, the seller knows whether there's a problem with the engine. They know whether it's been making a sound every time they drive it. So the seller knows the condition of the car.

And that's probably relevant, actually, to both of your valuations. If it's a high quality car, the seller probably values it more because they could sell it to someone else. The buyer probably values it more as well. I'd argue the buyer probably also has some private information. What does the buyer know? And these are things-- when I say the seller knows this, but the buyer doesn't.

And of course, these things aren't binary. The buyer probably knows something about the car. They can look at it and see if it's totally dented. They can see that, but there's probably some aspects of the car that they can't see or can't observe. What about the buyer? What is something the buyer may know? Yeah.

**AUDIENCE:** The buyer knows how much they want a car. I feel like that informs how much they'd be willing to pay for it.

**IAN BALL:** Exactly. So the buyer knows-- and this is exactly what you mentioned in the front row-- they know their valuation for the car. Buyer knows, or at least they know-- maybe they don't know their valuation exactly, but they know-- maybe I'll say how much they need it. The reason I don't say they don't know their valuation is if they don't know if the car is broken or not, then they probably don't know exactly how much it's really worth to them. But they know they have a job they need to get to tomorrow, and they have to get there.

And they know they went to the last three dealerships, and they couldn't find a car they liked. Or maybe they know that they like the color red, and this car is red-- or they hate the color brown, and that's why they value it less. There's a lot of aspects of the buyer's preferences that the buyer knows, but the seller doesn't.

And notice how this is fundamentally different from the imperfect information we see in poker. In poker, yes, I don't know if you have queens or kings. But when you sit down at the table, we basically all know the game we're playing and we all know the preferences of the other players. Another great example would be health insurance-- so also buying health insurance.

I would argue in this case probably the seller and the buyer know some things that the other party doesn't know very well. Any thoughts here? Certainly the buyer probably knows some things the seller doesn't know. Yeah.

**AUDIENCE:** The buyer knows his own medical history.

**IAN BALL:** Exactly. So the buyer knows medical history. And again, this is not binary. If you buy health insurance in the private market, the seller is concerned about this. And they're going to have you fill out a form. And they're going to have a doctor go to your house and do a 30-minute evaluation. But of course, that's not going to cover everything.

You might know that last week, you started feeling a bit sick, and that's why you're buying the health insurance. Or you might know about a family history of a condition that the health insurer doesn't know about. And this is certainly relevant to both of you. It's relevant to how much you value insurance, because the sicker you are, the more valuable insurance is to you. It's also relevant to how costly the insurance is to the seller. If the seller insures a very sick person, they're going to have to be on the hook for a lot of different doctor's bills.

And this is one reason health insurance markets often don't run very well. I'd argue the seller probably, even in this case, maybe knows some things. This is not the main missing information in health insurance markets, but I think there's probably some things they know.

**AUDIENCE:** Could the seller know that their company is about to do a discount next month? Does it tell the buyer?

**IAN BALL:** Great. That's a good thing. So one thing-- that's in a dynamic context. But certainly in a dynamic context, the sellers know the prices next period. This is certainly relevant. These clothing stores always do these regular discounts. If you go and buy something-- if you knew that they were going to have a 15% discount the next day, you probably wouldn't buy. And they usually don't tell you that the discount is coming in the future. They announced it when it started. They don't tell you in advance. So the seller might know things about future pricing.

And maybe this is a bit more of a behavioral, non-rational story, but I would argue that when you buy health insurance, it's very, very complicated. The exact quality of the network you have is very, very difficult to untangle. So technically, when you buy a health insurance plan, what's relevant to you is every single doctor in the world who's covered by the plan and who's not. And more precisely, exactly how much your copay is at every single doctor maybe in the world. You probably don't know that that well, even if you pore through the contract.

The seller probably has a much better idea of, maybe what I'll say-- the network coverage. Or what they certainly know about is when you apply to get reimbursed for a treatment, how cumbersome that process is, and what the rejection rate is of the insurer. This is something the insurer knows. Some insurers might be a bit more lenient than others. They privately know that. Maybe you've gone on Reddit and read a bit about it, but you probably don't have full information about that.

Let's do a few more examples. What about in an auction? Context? Well, here it's the classic case. When the bidders show up to an auction, the bidders generally know how much they value the item being auctioned, but the other bidders don't know that. So each bidder knows her value for the good.

And of course, this value could reflect a few things. If I'm bidding against someone in auction, their value might be high because they just love this painting or maybe they're just really, really rich. Or maybe they just got all this money and they have to spend it today. There's all these stories that could go on, but their willingness to pay for the item is something that they know that I don't know. And then a classic case that we've also discussed so far is maybe bargaining negotiation more broadly.

Let's say we gave the example of bargaining over a peace treaty. When both sides sit down at the table, they don't really know what terms of the treaty are acceptable. If I'm negotiating with a foreign leader-- I would never be doing that-- they know their political constraints at home. They know what kind of deal they can make that would get them in trouble or not.

They might know facts on the ground. They might know things about the condition of their soldiers and their army that are all relevant to the negotiation that I, as the counterparty, don't know. So there's just tons of private information. I'll just say about really, what is it about? It's about their preferences over the terms of the negotiation, over terms of the deal.

So I think it's nice to step back here and pause and look at the history of game theory. Game theory started-- as I said earlier in the course, it was studied by mathematicians. They were analyzing things like chess and checkers and poker, and they presented this theory of complete information that did a very good job of capturing games like chess and checkers and poker.

But there was this folk wisdom. There was this view that we could never apply game theory to the real world, to things like this because the game isn't common knowledge. There's just so much information that we don't know that we can never apply the tools and the rigor of game theoretic reasoning. And there was a big breakthrough in the '60s.

So the breakthrough came with Harsanyi in 1967, 1968, who presented a framework for analyzing situations like this using game theory. And I think this was a turning point in game theory, where before that, it was largely studied by mathematicians. And then once people realized that this is the right way to think about health insurance markets or bargaining, all of a sudden economists became much more interested in it, and the nature of the field kind of shifted. And Harsanyi later won the Nobel Prize for his work here.

So I think Harsanyi's key insight, which was later formalized by others, is that we can model incomplete information like we model imperfect information. So in other words, I think as a sentence, this doesn't really make much sense, so let's say what we mean. What Harsanyi said is maybe we can model negotiation and health insurance the way we model poker. So why don't we think about someone's health status in healthcare or a military leader's private information about their army the way we model your hand in a game of poker?

So the key is that we don't literally have a dealer who's dealing you cards, but we can model this private information as the move of a metaphorical dealer, or metaphorical move by nature who deals people their private information. So Sonia's idea is that we can have a metaphorical first move by nature. And I want you to think about this.

I think the best way to think about this is nature is playing the role of the card dealer. But what they're dealing you is your health status or how much you value the painting or the condition of your army. And after nature makes that deal, then we can analyze the game like a game of imperfect information.

Now, I think a challenge here is that this is metaphorical. So I think in poker, it's very clear how we can reason through a situation before the cards are dealt. If I'm playing poker, I'm sitting at the table, I don't have the cards yet. The dealer deals me a card. I know that's going to happen. And I can think about when that's going on.

It's a bit hard to think about you buying health insurance at the stage before nature dealt you your health status. There is this subtlety in how we interpret it, but despite the difficulty of interpretation-- so maybe I'll say this creates some interpretational difficulties. And we'll see those throughout the course.

But nevertheless, we can analyze these things mathematically and we can make some progress. One difficulty, I guess, in the first group by nature is we have to specify the probability distribution over what nature chooses. That's a lot easier when nature is drawing a card from a deck. We all understand that. When nature is drawing your health status or your valuation for a good, that becomes a bit more challenging. So one challenge is that we have to specify the distribution over nature's moves. But let's see if we can get into this and start analyzing this.

So I could present the whole general model, but I think it's very complicated at first. And you might say, why do we need this complicated model? So let's start with an example to try to motivate why things are subtle and why we actually need to do some work to model this.

So let's think about a classic example. Let's say we're working on a group project. And let's say we have two players. And I'll call them  $i$  equals 1 and  $i$  equals 2. And each player,  $i$ , is going to choose effort on the project. I'm going to choose effort  $e_i$ . The player  $i$ , they choose effort  $e_i$ . And let's say that's a number between 0 and 1. So 0 is they don't work at all, 1 is they work as hard as they possibly can.

And now we're going to say the payoffs to player  $i$ -- well, player  $i$ 's payoff depends on how well the project turns out minus the cost they have to bear to exert effort. Let me see if I had  $1/2$  in here. So let's say that their payoff is  $e_1$  times  $e_2$  minus  $e_i$  squared.

So what this says is  $e_1$  times  $e_2$ -- the product of the effort that the two players put in, this is going to represent basically how good the project is, how good the output is. And this is the cost of effort. So notice, if I'm player  $i$ , my effort  $e_i$  affects my cost. And it also affects the quality of the output. The other player's effort affects the quality of the output, but not my cost, because I don't bear the cost for the effort that the other player is putting in.

This is a pretty simple game to analyze, but the trick is we're now going to put a  $\theta$  up here, where  $\theta$  is going to be a state that represents how hard the project is. When you get working on something together, you often don't know how hard this task is. So  $\theta$  is going to represent the difficulty of the project. Actually, a higher  $\theta$  means the project is easier. So maybe I'll think of  $\theta$  as how easy the project is.

So as a simple case, we could say maybe two values,  $\theta_H$  and  $\theta_L$ . So these are the two possible values  $\theta$  could take.  $\theta_H$  is a low value because if the project is hard, when you put in this much effort, you don't get very much output. If the project is easy,  $\theta_L$ -- and I guess I'm so used to high and low. So easy--  $\theta_H$ ,  $\theta_L$ . If the project is easy, then you get more output for the same level of effort. The problem is you probably don't know how hard the project is.

Let's think through a few cases here. So the easy case, case one would be no one knows  $\theta$ . If no one knows  $\theta$ , and they just believe  $\theta$  follows some distribution, then we can analyze this as a classic game, because at the beginning of the game, information is symmetric.

So here, this would be symmetric info. If no one knows  $\theta$ , then there's no information that one player knows that another player doesn't know. But the harder case-- let's look at case 2 where only player one knows  $\theta$ .

So if only player 1 knows  $\theta$ , let's see if we can start analyzing this game. Well, in case two, player 1 is going to possibly make a different choice depending on how hard the project is. So player 1 is going to make maybe one choice,  $e_1$  of  $\theta = H$ , and a different choice,  $e_2$  of  $\theta = L$ . If player 1 knows how hard the project is, if the project is hard, they might find it not as worthwhile to work on the project. This should be  $E_1$ , and this should be easy, so everything's wrong here.

If they know the project's easy, they might choose a different level of effort. So we're going to have to capture that when we think about a player's strategy in the game. Player 1's strategy might depend on how hard the project is. But now let's look at player 2. What should player 2 do?

Let's imagine you are in this situation. You don't know how hard the project is, and you're thinking about how hard to work. I don't know. I think there's some information that's missing here. There's something you need to know. So we've said they don't know how hard the project is.

So they don't know  $\theta$ . But I think we're missing some information to see how they should behave. I don't know  $\theta$ . I don't know. How would you approach this if you were player 2? Yeah.

**AUDIENCE:** Maybe in terms of a security strategy.

**IAN BALL:** OK.

**AUDIENCE:** Let's try to minimize the possibility that it goes wrong, given that-- I don't know if it's a-- I'm not sure.

**IAN BALL:** So that would be one way. No, I think that would be a reasonable answer. We're not going to pursue that too much in this course, but that's something that people do. And that's one thing you could study. Yeah?

**AUDIENCE:** If you know how much player 1 cares about the project, you could use backwards induction. If player 1 knows  $\theta$  and player 1 cares a lot, you will know that player 1 is just going to do the projects, and you don't have to do anything.

**IAN BALL:** So I think the first thing you said there was really crucial. You said, if player 1 knows  $\theta$ , and then you reason from that. But does player 2 know that player 1 knows  $\theta$ ? Well, we need to make a choice about that. So this is the crucial thing. It's not just enough to say, I've said player 2 doesn't know whether project is hard. But you immediately gave it away.

When you're reasoning about this, I need to think through how hard is player 1 going to work? But how hard player 1 works is going to depend on whether player 1 knows how hard the project is. So I need to think. Here's a question for the analysis. Does player 2 know whether player one knows  $\theta$ ? Because that's going to affect the way that player 1 moves, and therefore, affect the best thing for player 2.

But now I think we're getting into a regress. Let's say you're player 1. How are you thinking through this? What's relevant now for player 1 when they're trying to think? We know player 1 knows  $\theta$ . But player 1 needs to think, well, how hard is player 2 going to work? But how hard player 2 works depends on whether player 2 knows whether player 1 knows  $\theta$ . So now we have to think from player 1's perspective, does player 1 know whether player 2 knows whether player 1 knows  $\theta$ ?

And I think you can see where this is going, right? We care about what the other player is going to do. What they do depends on their beliefs about what I know, but also their beliefs about what I know that they know and my beliefs about what they know that I know that they know, and it unravels here. So we need a way to get a handle on this.

And I think this was what was seen as the key challenge to these games. At first, you might think, all we need to specify is what each player knows about  $\theta$ , about the state of the world, how hard the project is. But we soon realize that to fully analyze the game, we need to take a stance on what each player knows about what other people know, and what other people know about what other people know, and so on.

And so the key takeaway here is that we really need to distinguish two different aspects of the game. So I think this motivates a key distinction. I need another arrow here. Between the physical state  $\theta$ , this is, in this example, how hard the project is. It's totally outside the minds of the players. It's either hard or it's easy. That's just some fact of the world, fact of nature.

But then we also need to think about maybe the mental state of each player, which maybe before we saw this example, we might have thought, oh, mental state just means their belief about  $\theta$ . But now it really means their belief about the mental state of the other player. And in general, we're going to try to have this mental state capture all the relevant beliefs about the physical state and also about the mental state of the other player. And that's the direction we're going to go now with this formal setup. Any questions about this? So I don't expect people to see the solution, but just to see that there's a problem is the goal of this example.

So now we're going to formally present our model of a game that can capture incomplete information. And we're going to be using Bayes' rule a lot, so these games are often called Bayesian games. So how do we try to capture a situation like this? Well, first we're going to have the players.  $i$  equals 1 up to  $n$ . And then we're going to have the actions. So each player-- so player  $i$  is going to choose some action in some action set.

So just like in repeated games, if we were to call this a strategy set, we would then run into issues, because in the meta-Bayesian game, there's going to be a different notion of strategy. So I want to use the term action instead of strategy. I think the notes talk about Bayesian games with Bayesian extensive form games at this part. We're not going to talk about that, at least for now.

So we're going to think of this as just a simultaneous move game. We're each choosing an action, just like we did over there. We each choose how hard to work. But then there's going to be some state of nature, maybe our physical state or sometimes called a state of nature.  $\theta$ , which is going to be in some set, Big  $\Theta$ .

I don't know if we've used  $\theta$  before. When I write  $\theta$ , it just looks like an  $H$  inside a circle. But that's so we can see the difference. Little  $\theta$  is like this, big  $\theta$  is like this. So I know it looks weird, but that's how I write it on the board. And just so we understand, how we map it into our example over there-- there we had two players. Each player's action was some effort level between 0 and 1.

And the physical state was the difficulty of the project, which over there, there were two states. It could either be a difficult project; H, hard project; or an easy project, theta e-- an easy project. And then we're going to need payoffs, just like we had over there. So then we're going to have utility for player i. This is going to depend on the actions that every player takes, but also potentially on the state of the world.

So just like over here, my payoff depended on my own effort level, my opponent's effort level, and also the state of the project, how hard the project was. So over here, in the more abstract context,  $U_i$  is going to be a function of  $A_1$  all the way up to  $A_n$ , the actions of every player, times the set capital theta. And it's going to go to R for each L. So this is going to give me a real numbered payoff that depends on what player 1 does, what player 2 does, all the way up to what player n does, and also the value of the physical state-- in this example, whether the project is easy or hard.

But what we see from that game is we haven't specified enough yet because we also have to think about the mental states, what players know about other players. And Harsanyi's key trick is to introduce this abstract object, which we call a type. And the type of a player is going to capture their complete mental state. It could be their beliefs about the state of the world. It could be their beliefs about the other player. It could be all these things.

So we're going to say player i has a type,  $t_i$ , which is in some set, capital  $T_i$ . So this is called player i's type space. And type is a weird terminology, but you can say there's the type of the player who knows the state. There's the type of the player who doesn't know the state. There's the type of the player who knows the other player knows the state, but doesn't know the state. We just think of these players with different mental states as different types of players.

And what we're missing here is what's the connection between the types and the state of the world. And we're going to capture that with a prior. So the last component here is going to be a distribution, P. But it's not enough just to say the distribution of the physical state, but we also have to specify the distribution of the players types. And more generally, we have to specify the distribution over the combination of natural states and types.

So this distribution, P, is going to be over theta cross  $T_1$  cross  $t_n$ . So it's going to tell us, what is the probability that the physical state is this? Player one's mental state is something, all the way up to player N's mental state being something over here. I think this is a bit abstract, so let's try to go through our example. Maybe I'll give one.

Yeah, so let's go try to capture this in our example, where I didn't specify everything. Let's say, does player 2 know whether player 1 knows theta? We're going to say for this example yes. And does player 1 know whether player 2 knows whether player 1 knows theta. We're going to say yes. So we could make it more complicated, but we're going to say all these things are known.

So in this example, player 1 knows the state theta. Player 2 doesn't know the state theta, but they do know that player 1 knows the state theta. And we're going to try to capture that in our framework over here. So the first, I think, challenging step whenever we set up these models, the first thing is to specify the state. That's easy-- the state theta. And then we have to specify the type spaces, and that's always a bit harder.

So let's go through this example. What is the safe space  $\theta$  going to be if we want to capture our example of the group project over there? What were the physical states in that example? Yeah. Right. So we just have-- and actually,  $\theta_H$  and  $\theta_E$ . I wrote it wrong. So often we do high and low, but here, let's think of it the project is hard or the project is easy. These are the two physical states.

Did I write-- I don't know. I did originally make a mistake, so that's on me. But now the hard thing is specifying the types. This is always the challenge. So the next step is to specify  $T_1$  and  $T_2$ . So I think the question you have to ask yourself is, how many distinct mental states does each player have?

Any thoughts? Let's think about player 1. So we know that player 1-- player 1 knows the state. So what are some distinct mental states that player 1 could have? Yeah.

**AUDIENCE:** Player 2 [INAUDIBLE]. Player 1 can know about player 2 but doesn't know that player 1 knows [INAUDIBLE].

**IAN BALL:** So there's a lot of things that could happen, and this is why I tried to simplify it by saying we're going to assume that player 1 does know whether player two knows the state. So you're right. If we really wanted to get general, we'd have to deal with that. But we're going to assume player 1-- these higher order beliefs are known. We're going to make it as simple as possible.

So player 1 does know all these things, but I would still argue there's certainly two different mental states player 1 could have. What are these two mental states? We said player one knows the state, but what does that mean? What could player 1 believe? If they know the state, well, either they know the state is easy or they know the state is hard.

So there's at least two mental states. There's the type of player 1 where they know this is a hard project, and there's the other type of player 1 where they know this is an easy project. So in this case, if we really went down this rabbit hole, we'd have to make it harder. But in this simplest example, there's actually only two types. And they exactly correspond to the mental state where player one sees or knows the project is easy and the mental state where they see or know the project is hard.

There's a pretty close correspondence here, but to keep the distinction clear, let's call this  $t_H$  and  $t_E$ . So the interpretation is if player 1's type is  $t_H$ , what does that mean? That means player 1 knows the project is hard. Notice that's different from saying the state is  $\theta_H$ . When I say the state is  $\theta_H$ , that's a statement about whether the project is actually hard or not. When I say that player 1's type is  $t_H$ , that's a statement about what player 1 knows or believes about the state.

What about player 2 here? It turns out in this really simple case, there's only one mental state for player 2, because they don't know whether the project is hard or easy. They do know that player 1 knows that, so they show up and they play the game. There's only one possible type, so maybe we'll just call this  $t_0$  to show They don't really know anything.

So we now have specified the physical states and the mental states. We could have made it much more complicated. If there was uncertainty about what the other players knew, we'd have to model larger type spaces. But for the simple example, we've made it quite nice. And let's also assume that  $\theta_H$  and  $\theta_L$  and  $\theta_E$  are equally likely.

And this is known among the players. So to be clear, player 2 knows that it's equally likely that the project is hard or easy. They know that player 1 knows that the project is hard or knows that it's easy, but they don't know anything else. So now the challenge is, can we describe this distribution,  $P$ ? So this distribution,  $P$ , is going to be a distribution over  $\theta$  cross  $T_1$  cross  $T_2$ .

So what we need to specify is four different probabilities. So we need to say, what's the probability of  $\theta H, tH, t_0$ ;  $\theta H, tL, t_0$ ;  $\theta E, tH, t_0$ ;  $\theta E, tL, t_0$ . So sometimes these are called states of the world. The distinction is  $\theta$  is a state of nature or physical state. The state of the whole world specifies both the physical state and the mental states.

So these are the four possible states of the world, and let's interpret them. Let's look at the second one-- just an example. This is the state of the world where the state of nature is  $\theta H$ . So the project is hard. The player one's mental state is  $tE$ . That means player 1 believes that the project is easy. And player two's mental state is just  $t_0$ . They believe what they always believe. They don't really have anything to believe. They just have the one thing they know.

So what we want to do is we need to assign probabilities to each of these four things based on the story we've told so far. So any thoughts on what this probability distribution,  $P$ , is going to be here? Yes, back here.

**AUDIENCE:** Would it be  $1/2, 0, 0, 1/2$ ?

**IAN BALL:** Exactly right. So can you explain a little bit how you got that?

**AUDIENCE:** For sure. So if the state of the world  $\theta H$  is-- let's just say  $\theta$  is  $\theta H$ . That means we know--

**IAN BALL:** State of nature. Sorry, it's very confusing. Yeah, if the state of nature is  $\theta H$ . Yeah, exactly. I may, have said it wrong myself. Yeah.

**AUDIENCE:** Anyway, then, with absolute certainty, we know that player 1 knows that the state of nature is  $\theta H$ , which means they have  $tH$ .

**IAN BALL:** Exactly.

**AUDIENCE:** It's not possible for them to have  $t$  and it's this mismatch, which is why that gives probability 0. And then you flip the labels and everything else is symmetric.

**IAN BALL:** Exactly. Exactly right. I guess it's a hidden implicit assumption that when we say player 1-- well, this is what it means-- to know the state means when you think the state is high, it's actually high or hard. And when you think the state is easy, it's actually easy. Otherwise, you wouldn't know the state at all. OK, let's do maybe one more example to make sure we can understand what's going on here.

So let's consider maybe example two, where we'll try to be a little more realistic. Probably player 1 can't perfectly tell whether the project is easy or hard, but they have some information about it. So what we're going to imagine is player 1 gets a signal about the difficulty of the state. And so player 1 has some information about whether the project is easy or hard, but they don't know it for sure.

And more specifically-- so let's say player 1-- sometimes we say gets a signal. And that's a statistical way of looking at it, but you sit down, you try to start doing your homework. You think, I think it's probably hard, but I'm not sure. And you maybe assign some probability to how hard it is. And that's what we're going to try to capture here. I'm going to say with accuracy,  $q$ , where  $q$  is going to be between  $1/2$  and  $1$ . So what this means is-- let me just draw this mapping.

I run out of arrows here. So what I'm representing here is I'm going to think of  $\theta_H$  and  $\theta_L$  as the true difficulty of the project. And I'm going to think of  $t_H$  and  $t_L$  as the signal that player 1 gets about the difficulty of the project. And if the project is hard, then most of the time, player 1 gets this signal that's indicative of it being hard-- indeed, with probability  $q$ . But with probability  $1 - q$ , it's deceptively hard. It's actually hard, but it looks easy.

And they get the wrong signal. And then we have the same thing over here. So let's just understand-- if I allow  $q$  to be  $1$ , then if the project is hard, they always get the signal that it's hard. And I'm just going to say all day long, they get it wrong.

Similarly, when the project is easy, they always get this signal that suggests it's easy. So the case of  $q$  equals  $1$  is exactly what we already studied over here. But now we're going to generalize to this case, what if  $q$  was  $1/2$ ? What would that correspond to? Yeah.

**AUDIENCE:** You're right half the time.

**IAN BALL:** Right. And if the prior-- since we said it's equally likely to be hard and easy, it actually means you know nothing. If you have a signal when the state is hard, it gives you each signal realization with probability  $1/2$ . And when it's easy, you get each signal with probability  $1/2$ , then it's not giving you any information at all. So you're just totally uninformed. Your gut is totally useless, which is maybe for retail stock traders. That's a pretty good model of how they trade stocks.

So let's now try to capture this model, this example within our model over here. So it turns out the type space are going to be the same, but the interpretation is different. So we still have, for player 1, two types-- the hard type and the easy type. But now the hard type, we don't mean you know for certain the project is hard. What we mean is based on the evidence you see, it's more likely than not that the project is hard.

And similarly, the type  $t_E$  suggests that player 1 looks at the project, they think it's probably easy, but they're not certain that it's easy. And they can actually compute the probability that it's easy. So now let's represent this. We're going to represent this with the distribution,  $P$ , over  $\theta \times t_1 \times t_2$ .

So what is it going to be? Let's write it down. Again, we have  $\theta_H$ ,  $t_H$ ,  $t_0$ . We have  $\theta_H$ ,  $t_E$ ,  $t_0$ . We have  $\theta_E$ ,  $t_H$ ,  $t_0$ , and we have  $\theta_E$ ,  $t_E$ ,  $t_0$ . I guess it got progressively smaller as I ran out of space. So now what are the probabilities that we're going to assign to each of these? Yeah.

**AUDIENCE:** Just to make sure there is equal probability, that it actually is hard.

**IAN BALL:** We're going to maintain that assumption, so a good question. Let's maintain this assumption. I didn't say that. Second row, yeah.

**AUDIENCE:** So for the first one, there's a  $1/2$  chance that it's hard. And then as  $t_{dH}$ , that's going to be  $q$ .

**IAN BALL:** So what's the probability of this?

**AUDIENCE:** That's  $1/2 q$ .

**IAN BALL:**  $1/2 q$ , exactly.

**AUDIENCE:** [INAUDIBLE]  $1$  minus  $q$ .

**IAN BALL:** Yes.

**AUDIENCE:** And then  $1/2$  minus  $q$  and then [INAUDIBLE].

**IAN BALL:** Great. So it's going to be like this, because what has to happen for this to be the state of the world? Well, first, the project has to be hard. And we know that happens with probability  $1/2$ . But further, not only is the project hard, but player 1 gets the signal that's indicative of it being hard.

And when the state's actually hard, how likely is that? That's probability  $q$ . And then down here, well, we're going to say, well, there's still a probability  $1/2$  that it's actually hard. But now what's the probability they get the easy signal? So they get misled, they get the signal that suggests the project's easy even when it's hard. That's still possible. It's less likely, but it's possible. That has probability  $1$  minus  $q$ , and we have to multiply that by  $1/2$ .

Notice here if we sum the first two, we get  $1/2$ , which makes sense because the sum of the first two is exactly the probability that the state is hard, because given the state is hard, we're summing over the additional types. Let's look at the extreme cases. If  $q$  goes to  $1$ , what happens here? Well, now we're just back over here. This is what we said. This is the special case where player 1 perfectly observes the difficulty of the project. What if  $q$  goes to  $1/2$ ? Yeah.

**AUDIENCE:**  $q$  goes to half, then each of the pairs, like the data pairs are approximately going to be  $1$ .

**IAN BALL:** So basically everything's a fourth because now your signal doesn't really give you any information about this. If the project is actually hard, you're equally likely to get these two signals. And just everything is going to be  $1/4$ . Yes.

**AUDIENCE:** Is player 1 aware of their own, or does that even matter?

**IAN BALL:** No, it really does matter. So this is the subtle thing here that this is where the implicit assumption here-- and this is where it gets a little subtle-- is that the distribution,  $P$ , is common knowledge. So player 1 knows their own type, and they know the set of possible types the players have. And they know the distribution over the types.

Now, you might say, is that a restriction? It turns out there's some really subtle mathematical points that argue that, as long as you make the type space big enough, the  $P$  is a little tricky. Basically, you can make the type space big enough and you can capture everything, but it's quite a subtle point about what's known and what's not known. And we're getting into this very abstract area of epistemic game theory that gets into logic and really subtleties about what belief and what knowledge is. But the assumption here is that  $P$  is known by everyone, yeah.

OK. So the next thing we want to do is let's now think about behavior in this game. So far, I've just defined the game. We now have to define-- the remaining time, we're going to define an equilibrium of this game.

So far, I've just defined what a Bayesian game is. Let's now define what an equilibrium of a Bayesian game is. The first step is to define equilibrium. You always have to define what a strategy is. So now let's define a strategy in a Bayesian game. Now a strategy, remember, is always a complete contingent plan. It specifies what action you take as a function of things that you can observe, things that you know.

So in this case, for player  $i$ , a strategy is a function,  $s_i$ , from  $t_i$  to  $A_i$ . So let's visualize it down here. We have some type,  $t_i$ , and that gets mapped to  $s_i$  of  $t_i$ . So what this complete contingent plan says is it says if I'm player  $i$ , if my mental state is  $t_i$ , this is the action I'm going to take.

If I have a different mental state, I can potentially take a different action. My strategy can't depend on the mental states of anyone else or on the state of the world or the state of nature, because I don't observe those things. All I know is player  $i$  is my mental state, so the only thing I can contingent condition my action on is my own mental state. And that's exactly why this function depends on my mental state, but on no one else's mental state.

I think there's one subtlety here in interpretation, which is that-- I'll say note, player  $i$  actually has one mental state slash type. So this is where I mentioned the interpretive difficulties we get into. I'm calling this a complete contingent plan, but it's never clear when you would make this complete contingent plan. You'd have to make this complete contingent plan before you know what your type is.

And it's very hard with these intrinsic preferences. You're born. Let's say you like the color red in our car example. Formally, what you'd have to do here is make a strategy that said, if I were born and I liked the color blue, what would I do? If I were born differently and liked the color red, what would I do? It's a bit weird to think about, but I think the role of this assumption really is that, yes, player  $i$  is only going to have a single realized type, and they're going to take a single action based on that type.

But the reason we have to specify this complete contingent plan is that the other players don't know what type player  $i$  has. So the other players are going to have to form beliefs about player  $i$ 's type, and then form beliefs about what actions player  $i$  takes. And the other players need to-- so what this represents is what the other players believe player  $i$  will do when player  $i$ 's type is  $t_i$ , even if player  $i$  wakes up and their type is not  $t_i$ . Maybe a little philosophical digression, but let's just stop there. Yes.

**AUDIENCE:** So does that mean like the strategy is basically information that everybody has? Everyone knows everyone else's strategy.

**IAN BALL:** So this is a subtlety. Normally in game theory, we make knowledge assumptions, not about what people do, but about what people know and about the circumstance. And then we define a solution concept, like an equilibrium. So yeah, do people know this is always a subtle thing? Do people know the strategies that are chosen in equilibrium?

I don't think we would formally say people know them. We would say that equilibrium specifies the beliefs that players have, and a demand of equilibrium is that those beliefs are correct. But we don't literally-- if we wanted to model the game where you actually observed players strategies, that would be a different game. We generally only model knowledge about exogenous things from nature.

We don't really think about knowledge about what other rational players do because that's just not a probabilistic process in the same way that's in their control. This gets into weird things about free will and all these things. But the point is, we don't assume knowledge of other people's strategies. In equilibrium, people's beliefs about other people's strategies will be correct. And that's very similar, but I would say slightly different.

**AUDIENCE:** So where does the part come in where other players have to understand, oh, if player  $i$  is born like red--

**IAN BALL:** When we define equilibrium. So I haven't defined equilibrium yet, so we'll see when we define equilibrium. That's great. That's a great question. OK, so the final step when to define equilibrium is you have to say a little bit more about beliefs. So let's put ourselves in the position of player  $i$ . Suppose I'm player  $i$ . What I know is I know my type.

So let's say my type is  $t_i$ . Now I have to decide how to play. I have to decide what action to take. But the action for me is going to depend on a lot of other things. So the first step is what I need to do is I need to form beliefs about  $\theta$  and about  $t_{-i}$ . I need to form beliefs about the state of nature  $\theta$  and about the types of other players.

In this model, why do I need to form beliefs about  $\theta$ ? Why do I care about  $\theta$ ? Well,  $\theta$  directly affects my payoffs. So going back to the example over here, I need to form beliefs about how hard the project is. Because if the project is really hard, I might not want to work very hard. It may not be worth it. The second, subtler thing is, why do I need to form beliefs about other players' mental states? Let's think of the project example. Why do I care what another player thinks about how hard the project is? Yeah, up front.

**AUDIENCE:** They'll probably adjust their effort for you.

**IAN BALL:** Exactly. So I care about their mental states, not because their mental states directly affect my payoff, but because their mental states will affect the actions that they take, and the actions that they take will affect my payoff. And this is where, going back to your question about the red-blue thing, I'm forming beliefs about their mental states, and then their strategy is going to specify what they actually do under those mental states. So it's that combination that's going to give me information. So the first thing is we have to learn how to use Bayes' rule. To do this, we use Bayes' rule.

We'll write something like this. What does this mean?  $P_i$  says I'm player  $i$ .  $t_i$  says this is my mental state,  $t_i$ . Given that this is my mental state, I want to say, what is the probability that the state of nature is  $\theta$  and the mental states of all my opponents are  $t_{-i}$ ?

OK. And this is going to be a classic application of Bayes' rule. I hope people have seen this before. If you haven't, it may be something you have to brush up on a little bit. But what is this? Well, the probability of  $A$  given  $B$  is the probability of  $A$  and  $B$ , divided by the probability of  $B$ . This is what Bayes' rule says.

And I'm going to go over it too fast, so if you haven't seen it, this won't make sense. I think you'll have to study it on your own. I'm happy to talk after class. So what are we going to get? On the top, we're going to have the probability that my type is  $t_i$  and this is the state of the world, so we're going to get  $\theta, t_{-i}$ .

On the bottom, we want the probability of  $t_i$ . This is a little more subtle. It's going to be a sum over all maybe  $\theta', t_{-i}'$  of  $p(\theta', t_{-i}')$ . And in words, this is the probability of  $t_i$ .

So Bayes' rule is what allows me to leverage the prior distribution,  $P$ , which, as we said, everyone knows to update my beliefs and form my beliefs and assign probabilities to the state of nature and to other people's mental states as a function of my mental state. And in general, this probability here could vary with my mental state. If I had a different mental state, I would form different beliefs about the other players.

OK, so now I think we can finally write down the definition of equilibrium. So let's maybe go down here. It's a bit of a long definition, so I really just want to spend the last 15 minutes just writing down one equation to make sure we really understand it. So we have our Bayesian game over there. And here's our definition.

So a strategy profile,  $s^*$ , which equals  $s_1^* \dots s_n^*$  -- so as usual, throughout the course, whenever we define an equilibrium, the object that an equilibrium is is always a strategy profile. It's a specification of a strategy for each player. So a strategy profile is a -- well, what do we call this? Well, it's a Nash equilibrium in a Bayesian game. So we call that a Bayesian Nash equilibrium. We just put the words together. It's a Bayesian Nash equilibrium.

Well, what's the basic idea about equilibrium? In equilibrium, every player is choosing their action optimally given the beliefs they have about the other players and those beliefs are correct. So we want to make sure that everyone is behaving optimally, given the way everyone else is behaving. So I think the first step is to figure out what we need to quantify over.

So what we want to consider is we want to say for every player and for every mental state that player could have, they're behaving optimally. So if for every player  $i$  and for every type  $t_i$  in capital  $T_i$ , the following holds -- so just let's understand in words what we want to write here. What we want to capture down here is to say that when player  $i$  has mental state  $t_i$ , the action they're choosing is optimal given their beliefs. And because we're requiring that for every player  $i$  and for every type  $t_i$ , that's capturing what we want

So first, let's compute what is the expected payoff of the player of type  $t_i$ ? So what we're going to do is we're going to take an expectation. And when I write this, I'm going to probably need a bit more space. I'm going to put something in here. But what I'm writing when I write the expectation given  $t_i$  is I'm saying, given  $t_i$ , I can compute conditional probabilities of these other things. And what I'm going to put in here is going to be a function of these things.

And I'm computing the conditional expectation of that quantity given my type,  $t_i$ . So what I want to put in here is my payoff when we all follow these strategies. So it's going to be my payoff. I'm player  $i$ . Now, my payoff depends on the actions that we all take and the state of the world, the state of nature.

So my player 1's action is going to be what? Well, player 1's action is going to be  $s_1$  of  $t_1$ . And this is exactly reflecting the point we made. Why do I, as player  $i$ , care about player 1's mental state, about their  $t_1$ ? Because their  $t_1$  affects the action that they're taking under this equilibrium strategy profile. And then it's going to go all the way up to  $s_n$  of  $t_n$ .

Now, it's tricky. One of these things -- the  $i$  position here is actually something I'm controlling as player  $i$ . But for everyone other than player  $i$ , these are random things that I'm forming beliefs about. So here, we have the action profile. And then the final thing is the  $\theta$ .

So what you can think of in here is if the state of the world-- not the state of nature, but if the state of the world is  $\theta_1$  all the way up to  $\theta_n$ , then under the strategy profile, these are the actions that all the players are going to take. This is the state of nature. And therefore, player  $i$ 's payoff is this. And player  $i$  is simply computing expectations over that. Questions about that?

So we want to compare that to what happens if player  $i$  deviates. So what I want is I want to now compute the same expected payoff if player  $i$  chooses some action-- let's say  $A_i'$ . And we're going to want this to be true for all actions  $A_i'$ . So it's also going to be a conditional expectation given  $\theta_i$ . And it's a little tricky that the  $i$  component here is hidden amongst the rest, so I want to reorganize this a little bit.

So another way of writing this which is useful is to write this as  $s_i^*(\theta_i, s_{-i}^*)$ . This is just notation where I'm going to put player  $i$  first and put everyone else's grouped together over here. This is really, I think, misleading notation when I write  $s_{-i}^*$ , because really, this is a collection of functions, each of which depends on a single player's type.

So this is  $s_2^*$ . Let's say player  $i$  is 1. This is  $s_2^*$  of  $\theta_2$ ,  $s_3^*$  of  $\theta_3$ ,  $s_4^*$  of  $\theta_4$ , and so on. It's a little subtle, but that's what's going on here.

And then what do I have over here? Well, I'm player  $i$ . I'm unilaterally deviating, so instead of taking action  $s_i^*(\theta_i)$ , which I'm supposed to take, I'm instead choosing action  $A_i'$ . So we have  $A_i'$  here, but everyone else's action is still the same. So they're still having  $s_{-i}^*(\theta_{-i})$ , but then my payoff also depends on the state.

So we have the state here. And this is our definition. And this is to return to this question of do I like blue or red? When I'm player  $i$ , I'm forming beliefs about the other player. And I might not know if they like blue or red. So I'm taking into account there's some probability they like blue, and this is the action they take. There's some probability they like red, this is the action they take. Even though when they wake up in the morning, they either like blue or red and they're going to take a single action. Yes.

**AUDIENCE:** Is there a reason why we have the condition that's given  $\theta_i$ , based on the fact that it can only technically have your  $s_i^*$  for a certain  $\theta_i$ ? And then you need to pin down your own mental state before you come up with a strategy?

**IAN BALL:** So I think there's two separate questions there. One question is, why is the domain of  $s_1^*$   $\theta_1$ ? And why do I have this conditioning here? So it's two questions. One is, why is a strategy a function from my type to my action? It's because the only thing I know is my type, so the only thing that my action can depend on is my own type. It can't depend on things I don't observe. Just like in poker, my strategy can't be bet if my hand is higher than the other person, because I don't know his hand.

So of course, that would be a great strategy. But it's not a legitimate strategy because I don't know what the other player's hand is. So that's why  $s_1^*$  is defined as a function of player 1's type only. Now we go here, and this is let's think about the decision stage. When I wake up and make a decision, I know something. I know my own mental state. If I'm player  $i$ , I know my mental state,  $\theta_i$ . And that gives me information. I'm going to update my beliefs based upon that.

So in the same way, going back to our example here, when player 1 knows the project is hard, they have a different belief about whether the project is hard than if they wake up and they know the project is easy. And they're going to behave differently in those situations. This is exactly reflecting that. It's reflecting the fact that my mental state controls what I believe about this theta. So just to understand, if we put in a different  $t_i$  here, this conditional expectation would change because the belief about theta and about everyone else's mental states could be different if this  $t_i$  were a different  $t_i$ -- if it were  $t_i$  prime.

**AUDIENCE:**  $t_i$  and  $t_i$  prime basically lead to different vectors of probabilities for other people's mental states as well as theta.

**IAN BALL:** Exactly right. So remember we wrote down this formula here. This is really a whole vector because once I fix  $t_i$ , I can vary the theta and the  $t$  negative  $i$ . So it's really a vector of probabilities. It's a lottery over everything I don't know, over all the theta  $t$  negative  $i$  pairs.

**AUDIENCE:** So this is kind of similar to how we talked about belief.

**IAN BALL:** No, it's exactly a belief. What we're talking about precisely is given my  $t_i$ , I form beliefs. The tricky thing is what I form beliefs about is  $t_1$  to  $t_n$  and theta. My belief about  $t_1$  to  $t_n$  by itself isn't that useful, but when combined with the strategy the opponents and the other players use, now it becomes useful.

Just knowing that the other player's mental state is  $t_2$  prime doesn't tell me much. But in equilibrium, I correctly anticipate that they're using the strategy  $s_2$  star, so I can anticipate that the action they take is  $s_2$  star of  $t_2$  double prime. And then that gives me information. Any other questions?

**AUDIENCE:** Yeah, I think you did just say.

**IAN BALL:** No, no. This is good.

**AUDIENCE:** [INAUDIBLE] another player has, for example, belief or not belief type  $t_2$ , then they're going to do this action. We can't know that for sure.

**IAN BALL:** So this is getting into, I guess, the notion of equilibrium. I think the short answer is yes if you want to get the problem right, if you want to be really subtle about what knowledge means. I guess one way of calling this a Bayes Nash equilibrium is if we all believe that these are the strategy profiles we're employing, then no one has an incentive to deviate from the way they're behaving.

So it's capturing the idea of Nash equilibrium that I'm believing. So just like in Nash-- in Nash, we say, I'm believing my opponents are following this strategy. And given that belief, I'm behaving optimally. If I believe that my opponents all follow this strategy profiles, then it's optimal for me to follow my strategy profile. And then that works for everyone, so we have this consistency. But I guess I would say there's a bit of a distinction between maybe this is a point.

Given  $t_i$ , I form beliefs about  $t_1$  to  $t_n$  and theta. These are totally outside the control of people. This is just something about the world. I can form those beliefs. Then the mapping from their types to the actions they actually take-- that's something that's under control of a rational decision maker. And we view that differently.

So generally, I guess this gets to a broader point of game theory. We view a fundamental difference between when I form beliefs about whether a dealt hand is going to be kings or aces and whether my opponent is going to bet a certain amount, because one is the choice of a rational player and the other is just something that's chosen by nature. So we maintain a distinction there in game theory.

**AUDIENCE:** You know how another player will play given a certain hand, but you don't know whether they actually have that hand.

**IAN BALL:** Yeah, I would say you have beliefs about the hand they will have. And then in equilibrium, the equilibrium describes consistent beliefs about the way they're going to play. But I guess maybe another way of saying it is if I change the strategies-- so yeah, this is maybe the key point. This is independent of strategies.

The probability I assign to the other player having this type is totally independent of how they play. This is not a statement about their behavior. This is just a statement about the world. Equilibrium is now a statement about behavior. But there's this fundamental difference there. But these are subtle things, so these are great questions to be asking. Yes.

**AUDIENCE:** So that just means-- so here, this is just over other people's beliefs. But then here, it's like what you actually think of what other people are thinking.

**IAN BALL:** Exactly. So by definition of my mental state, my mental state defines the beliefs I have about other people's mental states. But that alone doesn't allow me to form beliefs about other people's actions, because in general, I don't know the mapping from their mental state to their actions. We've defined equilibrium to be a profile of strategies that allows me to link my beliefs about their mental states to my beliefs about their actions, and we apply the classical idea of Nash equilibrium that says that this method of behavior is stable and consistent, in the sense that if everyone correctly believes that everyone's playing this way, then no one's going to want to deviate from this method of playing.

**AUDIENCE:** So this is based on what you think other people are going to play, but you already based that on what you think their beliefs will be.

**IAN BALL:** Yeah. I would say-- yeah, I think the short answer is this calculation is not an equilibrium object. This is a calculation that's a property of the game itself. This calculation is a property of the strategies and of the equilibrium concept. Maybe I'll just stop there. And let me say one more thing that maybe is useful, to see how we can tie this in to our notion of Nash equilibrium.

Actually, one thing I should have said before-- when we're computing these conditional expectations, you get into some weird issues when you condition on 0 probability events. So what I've assumed throughout, and maybe I'll say it now-- I should have said it before. Assume that this distribution,  $P$ , is strictly positive. So it assigns positive probability to everything, every combination of states and types. Say again?

**AUDIENCE:** The simple case in the beginning.

**IAN BALL:** Good, OK. So I was going to try to avoid that simplicity. But no, no. So let me say no. So when I say strictly positive, really what I mean is-- you're exactly right. Yeah. I should do it. I was going to skip a step, but you're right. It's good not to.

So this is really what I mean. It's not that every combination has positive probability, but every type of every player gets positive probability. And the reason I didn't want to write this out is what does this mean? I have to take a summation in order to compute this. This is a summation over all the other things. So it's true that there are some 0s here, but the probability that player 2 has type 0 is 1.

The probability that player 1 has type H is  $1/2$ , and the probability that the other player is type E is  $1/2$ . So that's really what I mean by positive probability. And that's actually about loss, because if something has 0 probability, I can just throw it out and not include it in the type space. So thanks for catching me on that. I tried to skip a step, but you kept me honest. Good.

And you'll see that's because when I'm conditioning, I'm conditioning on an event about my type alone. I'm not conditioning on richer things, and that's why I only need that to be positive probability. So if we want to understand what's going on here, one thing we could do is we could define a new utility function directly over strategies. So we could try to reduce this game and express it purely in strategic form. So we could say-- I'm going to write capital  $u_i$  of  $s_1$  up to  $s_N$ . And this is going to say, what is my ex ante expected utility if we all follow these strategies?

And now I'm going to take an expectation not conditional on  $t_i$ -- just an expectation before we learn any of our types. And it's going to be  $u_i$  of  $s_1$   $t_1$  and  $t_N$   $\theta$ , where here, the expectation is just with respect to this probability distribution,  $P$ , here, because the probability distribution,  $P$ , specifies the joint distribution of  $t_1$  up to  $t_N$  and  $\theta$ , so it makes sense to talk about this expectation. It's a bit of a weird object to think about because this is this weird calculation that you make before you're born and you know whether you like blue or red. It's a weird thing, but we can define this.

So now once we've defined this, we have a classic strategic form game. We can say what is the expected utility of each player,  $i$ , as a function of the strategies that are used? And now we have a theorem-- it's a pretty basic theorem. But we say that a strategy profile,  $s^*$ , equals  $s_1^*$  to  $s_N^*$ , is a Bayes Nash equilibrium of the Bayesian game if and only if  $s^*$  is a Nash equilibrium of the strategic form game defined here.

And I think this is a good homework exercise to try to go through and verify this. Yes.

**AUDIENCE:** [INAUDIBLE]

**IAN BALL:** All of these are implicitly defined over  $p$ . Maybe I'll put  $p$  here. But I guess the point I was trying to make is that this is a conditional expectation using conditional probabilities given  $t_i$ . This is an unconditional expectation using unconditional probabilities over  $P$ . That's the only point I wanted to make.

**AUDIENCE:** Based on the  $p$  in Bayes' rule.

**IAN BALL:** It's the same  $p$ , yeah. Sorry, this is the  $P$  based on Bayes' rule.  $P$  is a distribution. Using  $P$ , I can compute conditional probabilities here. Here, I don't even have to compute conditional probabilities because  $P$  is already a distribution over  $t_1$  up to  $t_N$  and  $\theta$ .

And maybe just two words about the intuition for this. What this says is whatever type I have, I'm behaving optimally. This is saying, let's imagine I'm in the position before I learn my type. it's saying that my complete contingent plan is optimal in expectation, but my complete contingent plan is only optimal if, at every realization of my type, it's optimal. And that's the connection, that if I anticipate that I'm going to make a mistake at some type realization, then there must be a mistake in my complete contingent plan.

And that's the connection here, but I think it's good to maybe write it out in algebra as an exercise. But let me stop there. I'm happy to stick around for questions, but let me conclude there. And I know this is abstract. It's natural for this not to make so much sense the first time, but I think doing problem sets will really reinforce this. Thanks.